A Publication of the American Psychological Association

Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Group Influence on Individual Risk Taking 75
Michael A. Wallach, Nathan Kogan, and Daryl J. Bem

Free Association Patterns as a Factor Relating to Semantic Differential Responses 87
Robert F. Terwilliger

Cognitive Balance and Self-Relevant Attitudes 95
Marc Pilisuk

The Effects of Reinforcement and Arousal on the Psychomotor Performance of Chronic Schizophrenics 104
Athan Karras

Attitude Change and Mental Hospital Experience 112
Jack M. Hicks and Fred E. Spaner

Epinephrine, Chlorpromazine, and Amusement 121
Stanley Schachter and Ladd Wheeler

Ordinal Position and Conformity 129
Selwyn W. Becker and Jean Carroll

CRITIQUE AND NOTES

A Developmental Study of Dogmatism during Adolescence with Reference to Sex Differences 132
C. C. Anderson

Prophecy Fails Again: A Report of a Failure to Replicate 136
Jane Allyn Hardyck and Marcus Braden

Emotional Arousal and Stimulus Duration as Determinants of Stimulus Selection 142
James M. Mattson and Thomas Natsoulas

Volume 65 AUGUST 1962 Number 2

ANNOUNCING
the publication of

AFFECT IMAGERY CONSCIOUSNESS

By **SILVAN S. TOMKINS**
Princeton University

This is the long-awaited full-length account—being published in three volumes—of the general theory of personality first presented at the 14th International Congress of Psychology, at Montreal, in 1954.

Silvan Tomkins says that psychology lost its heart and its mind from fear of methodological impurity, from excessive reliance on primary drives as motivators, and from attention to behavior rather than the complex transformations that make behavior possible.

He argues that psychology return to these critical problems abandoned to neurophysiology and biochemistry:

affect, imagery, and consciousness.

The human being is presented as an intercommunication system utilizing the feedback mechanism. Centrally emitted conscious messages are conceived as controlling both sensory and motor messages. Affect is represented as the prime motivating force—a biologically endowed response system which may instigate, support, modulate or reduce drives as well as other responses.

- **Volume I**
  The Positive Affects
  512 pages $8.00
  Published August 1962
- **Volume II**
  The Negative Affects
  about 600 pages
  (January 1963)
- **Volume III**
  Cognition and Ideology
  about 450 pages, with bibliography and index to the entire work (March 1963)

With the editorial assistance of Bertram P. Karon, Ph.D.

Orders for the set of three volumes, placed any time before publication of the third volume, will be billed at a subscription price that is 15% lower than the list price of the three volumes.

SPRINGER
Publishing Company, Inc.
44 East 23rd Street, New York 10, New York

Volume 65 August 1962 Number 2

Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

CONSULTING EDITORS

JOSEPH B. ADELSON
University of Michigan

GORDON W. ALLPORT
Harvard University

NORMAN H. ANDERSON
University of California, Los Angeles

LEONARD BERKOWITZ
University of Wisconsin

JACK W. BREHM
Duke University

URIE BRONFENBRENNER
Cornell University

DONALD R. BROWN
Bryn Mawr College

LAUNOR F. CARTER
System Development Corporation

DWIGHT W. CHAPMAN, JR.
Vassar College

MORTON DEUTSCH
Bell Telephone Laboratories

SEYMOUR FESHBACH
University of Pennsylvania

DONALD W. FISKE
University of Chicago

NATHANIEL L. GAGE
University of Illinois

NICHOLAS HOBBS
Peabody College for Teachers

MURRAY HORWITZ
New York University

IRVING L. JANIS
Yale University

HAROLD H. KELLEY
University of California, Los Angeles

HERBERT C. KELMAN
Harvard University

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

ROBERT B. MACLEOD
Cornell University

HAROLD L. RAUSH
National Institute of Mental Health

NEVITT SANFORD
Stanford University

IRVING SARNOFF
New York University

WILLIAM A. SCOTT
University of Colorado

JANET TAYLOR SPENCE
Veterans Administration Hospital, Iowa City

JOSEPH VEROFF
University of Michigan

JUSTIN L. WEISS
Massachusetts Mental Health Center

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued monthly, two volumes per year. The subscription price per volume is $10.00, or $20.00 annually (foreign, $20.50). Single issues are $2.00.

Original manuscripts should be sent to the Editor:

DANIEL KATZ
University of Michigan
Box 1248
Ann Arbor, Michigan

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR
Managing Editor

JUDY SMITH
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Lancaster, Pa., and at additional mailing offices

Copyright © 1962 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 75–86

GROUP INFLUENCE ON INDIVIDUAL RISK TAKING[1]

MICHAEL A. WALLACH[2]
Massachusetts Institute of Technology
NATHAN KOGAN
Educational Testing Service
AND DARYL J. BEM
University of Michigan

What are the effects of group interaction on risk and conservatism in decision making? By risk and conservatism we mean the extent to which the decision maker is willing to expose himself to possible failure in the pursuit of a desirable goal. Consider the situation in which several individuals working separately arrive at a series of decisions, and then are brought together to arrive at a group consensus regarding those decisions. What relationship should one expect to find between the individual decisions and the group consensus?

On the basis of prior experimental studies of individual and group judgment (e.g., Schachter, 1951; see also the section on group pressures and group standards in Cartwright & Zander, 1960, pp. 165–341), we should predict an averaging effect, i.e., group decisions randomly distributed around the average of the prediscussion individual decisions. Such an effect would seem to imply a process of minimizing individual losses, or minimizing the maximum individual concession. The cited studies report that inducements toward compromise and concession seem to be exerted most strongly toward group members whose initial individual views are most deviant from the central tendency.

An equally, if not more, compelling alternative hypothesis is that the group discussion will lead to increased conservatism, relative to the average of the prior individual decisions. One may cite the observations of Whyte (1956), among others, concerning the outcomes of conferences and meetings in bureaucratic organizations. Whyte argues that the use of committees and teams in the management of business and other kinds of enterprises leads inexorably to an inhibition of boldness and risk taking, a concentration on the conservative course when a choice must be made between more and less risky courses of action. How are such effects to be explained? First, it may be that the very nature of the group process or atmosphere encourages such a trend: there may be a fear, for example, of appearing foolhardy to others. Alternatively, or in addition, it is possible that the mechanism underlying an increase in conservatism is one of greater influence being exerted within the group by members whose individual conservatism tendencies are stronger. These two interpretations are not incompatible, of course, since the group process, if encouraging of conservatism, will enhance the influence of the initially more conservative members.

Finally, consideration should be given to the remaining and least likely possibility—that group interaction will eventuate in increased risk taking relative to the average of the prior decisions of the group members working separately. In this regard, Osborn (1957) has reported that group interaction may lead to quite radical, bold, problem solutions. While Osborn claims that special conditions must exist if such effects are to be observed, attempts to produce such conditions experimentally (Taylor, Berry, & Block, 1958) have yielded no evidence whatever for the so-called "brainstorming" phenom-

[1] This research was supported by a grant (G-17818) from the National Science Foundation. A master's thesis by J. A. F. Stoner at MIT's School of Industrial Management, with D. G. Marquis and M. A. Wallach as faculty advisers, was instrumental in inspiring the present investigation. We are greatly indebted to J. A. F. Stoner and D. G. Marquis for their aid and advice, and to V. Raimy and M. Wertheimer for facilities at the University of Colorado in Boulder. Thanks also are due S. Messick and A. Myers for comments.

[2] Now at Duke University.

enon. Thibaut and Kelley (1959, pp. 267–268) discuss the conflicting evidence on this issue. We might, in passing, also mention mass or crowd phenomena, in which extreme actions taken by groups are well beyond the capacities of the members of such groups considered individually (Brown, 1954; Turner & Killian, 1957). The relevance of such mass phenomena to group decision making in a laboratory context, however, is probably quite remote. In sum, increased risk taking as a consequence of group interaction appeared to us to be the least feasible of the three possibilities discussed above.

An examination of the literature reveals little experimental research which addresses itself explicitly to the problem of the present investigation. Lonergan and McClintock (1961) report that membership in an interdependent group led to no significant move toward greater conservatism or risk taking in a betting situation involving monetary gain or loss. Since the group situation was so structured that a consensus was not required, however, this experiment is not directly relevant to the aims of the present study. Hunt and Rowe (1960) report no difference between three-person groups and individuals in riskiness of investment decisions. However, the brevity of the group interaction (15 minutes) and the disruptive influence of having the various groups meet within sight of each other in a large room render their results inconclusive. Atthowe (1961), comparing individual and dyadic decisions in the choice of the better of two alternative wagers, found greater conservatism in the dyadic decisions. But the relevance of this result to the problem at hand is called into question when we learn that the alternative wagers were presented to the subjects as "problems taken from the mathematical reasoning section of an advanced intelligence test and arranged as wagers" (p. 115). This could well contribute to a conservative strategy.

We turn, finally, to a study by Stoner (1961), which provides the starting point for the research to be reported. Using male graduate students of industrial management as subjects, Stoner observed that a group consensus regarding degree of risk to be taken in resolving a "life dilemma" situation deviated from the average of prediscussion decisions in the direction of greater risk taking. These results took us by surprise. We wondered whether the finding could be generalized to other subject populations, whether it was an enduring effect, and whether it might have anything to do with relationships between risk taking and perceived group influence.

One issue that arises in interpreting Stoner's (1961) study concerns the effect that expectations about one's role might have on the results. Thus, a group of male graduate students of industrial management might make more risky decisions qua group than would each such student individually—the result obtained by Stoner—because the presence of their peers reminds each that one of the positively sanctioned attributes of the business manager role which they occupy or aspire to occupy is a willingness to take risks in their decision making. Stoner's use of a male business school sample, therefore, leaves open the possibility that his results may be a function of this particular group's self-assigned professional role alone. It also is possible that a group of males, regardless of their professional role, might make more risky decisions when gathered together because the presence of other males serves as a reminder that one of the expected indications of manliness in our society is a willingness to be bold and daring in decision making. Conversely, a group of females might make more conservative decisions when gathered together, or at least might fail to shift in a risky direction, since risk taking tendencies are not likely to be mutually reinforced in groups for whom risk is not a positive social value (see, e.g., Komarovsky, 1950; Milner, 1949; Wallach & Caron, 1959).

In the present experiment, we shall employ samples of male and female undergraduates enrolled in a liberal arts curriculum at a large state university. If the effects observed by Stoner (1961) are found to hold for both of the above samples, this would constitute strong evidence for the generality of the phenomenon and its independence of occupational and sex role considerations. Furthermore, the use of previously unacquainted subjects whose ascribed status is

initially equal will insure that whatever effects are obtained cannot be attributed to an association between initially high or low status, on the one hand, and risk or conservatism, on the other. If initial status levels were unequal, low status individuals might simply adopt the standards of those whose status is high—an outcome which would tell us nothing about the effect of group interactional processes as such on individual risk taking.

One should distinguish initially ascribed status from status indices (e.g., perceived influence and popularity) derived from the group experience. Since such indices may bear some relation to initial risk taking level, the necessary sociometric-type judgments will be obtained.

Finally, evidence will be presented with regard to the following two questions: Is the group induced effect on risk taking limited only to the group member's overt compliance in the group setting or does it also extend to his covert acceptance when he makes post-group decisions as an individual (see Festinger, 1953; Kelley & Thibaut, 1954)? To what extent are group effects on individual decision making relatively enduring or short-lived?

METHOD

Assessment of Level of Conservatism or Risk Taking

The instrument used for assessing level of conservatism or risk taking, as developed in some of our prior research (Kogan & Wallach, 1961; Wallach & Kogan, 1959, 1961), is called an "opinion questionnaire" and contains descriptions of 12 hypothetical situations. The central person in each situation must choose between two courses of action, one of which is more risky than the other but also more rewarding if successful. For each situation the subject must indicate the lowest probability of success he would accept before recommending that the potentially more rewarding alternative be chosen. The probabilities listed are 1, 3, 5, 7, and 9 chances of success in 10, plus a final category (scored as 10) in which the subject can refuse to recommend the risky alternative no matter how high its likelihood of success.

The situations were so designed as to cover a wide range of content, and may be summarized as follows:

1. An electrical engineer may stick with his present job at a modest but adequate salary, or may take a new job offering considerably more money but no long-term security.

2. A man with a severe heart ailment must seriously curtail his customary way of life if he does not undergo a delicate medical operation which might cure him completely or might prove fatal.

3. A man of moderate means may invest some money he recently inherited in secure "blue chip" low return securities or in more risky securities that offer the possibility of large gains.

4. A captain of a college football team, in the final seconds of a game with the college's traditional rival, may choose a play that is almost certain to produce a tie score, or a more risky play that would lead to sure victory if successful, sure defeat if not.

5. The president of an American corporation which is about to expand may build a new plant in the United States where returns on the investment would be moderate, or may decide to build in a foreign country with an unstable political history where, however, returns on the investment would be very high.

6. A college senior planning graduate work in chemistry may enter university X where, because of rigorous standards, only a fraction of the graduate students manage to receive the PhD, or may enter university Y which has a poorer reputation but where almost every graduate student receives the PhD.

7. A low ranked participant in a national chess tournament, playing an early match with the top-favored man, has the choice of attempting or not trying a deceptive but risky maneuver which might lead to quick victory if successful or almost certain defeat if it fails.

8. A college senior with considerable musical talent must choose between the secure course of going on to medical school and becoming a physician, or the risky course of embarking on the career of a concert pianist.

9. An American prisoner-of-war in World War II must choose between possible escape with the risk of execution if caught, or remaining in the camp where privations are severe.

10. A successful businessman with strong feelings of civic responsibility must decide whether or not to run for Congress on the ticket of a minority party whose campaign funds are limited.

11. A research physicist, just beginning a 5-year appointment at a university, may spend the time working on a series of short-term problems which he would be sure to solve but which would be of lesser importance, or on a very important but very difficult problem with the risk of nothing to show for his 5 years of effort.

12. An engaged couple must decide, in the face of recent arguments suggesting some sharp differences of opinion, whether or not to get married. Discussions with a marriage counselor indicate that a happy marriage, while possible, would not be assured.

The response categories are arrayed from chances of 1 in 10 upward for the odd items and in the reverse order for the even items, thus, counter-

balancing for any possible order preference effect in choice of probability levels. An overall conservatism-risk taking score is derived by adding the scores for the separate items. The larger this score, the greater the subject's conservatism.

Our prior research, cited above, yielded split-half Spearman-Brown reliability coefficients ranging from .53 to .80 for various age and sex samples, suggesting that the instrument possesses satisfactory internal consistency. The results of the present experiment will provide evidence, furthermore, of high test-retest reliability.

Regarding the instrument's construct validity as a risk taking measure, our earlier studies, cited above, have yielded findings consistent with a risk taking interpretation. For example, degree of conservatism as measured with the present instrument increases with age from young adulthood to old age for both males and females, and increases with degree of subjective probability of personal failure in a motor skill game with actual motor skill controlled.

Experimental Condition

Subjects. The subjects were invited to participate in an experiment which would take no longer than 2 hours and for which remuneration would be provided. Six subjects were scheduled for any one time, with every effort being made to insure that previously acquainted persons were not signed up for the same session. A total of 167 subjects participated in the experimental condition—14 all-male groups and 14 all-female groups.[3] The subjects were liberal arts students enrolled in summer session courses at the University of Colorado in Boulder.

Prediscussion individual decisions. The experiment was run in a seminar room around a very long table. For the initial administration of the questionnaire, subjects took alternate seats with the experimenter at one end. The six subjects were requested to read the instructions to the questionnaire and to look over the first item. The experimenter then emphasized two points in further standard instructions: that the more risky alternative is always assumed to be more desirable than the safer course, if the former should prove successful; that the odds which the subject marks indicate the lowest odds the subject would be willing to take and still advise the central figure to give the risky alternative a try. The subjects were told there was no time limit, that they should consider each of the 12 situations carefully, and that they could return to an earlier question if they wished to. The conservatism-risk instrument then was filled out individually by each of the six subjects in a group administration session that took about 20 minutes. To avoid giving any of the subjects the feeling that they were being rushed, the questionnaires were not collected until all had finished.

Group discussion and consensual group decisions. Without having had any prior expectation that they would be requested to discuss their decisions, the six subjects were then asked to move together into a discussion group at one end of the table. They now each were given another copy of the questionnaire, and a stand-up cardboard placard with the identification letter K, L, M, N, O, or P on it was placed before each subject. The experimenter then told them that the questionnaire now before them was the same one they just finished taking. They had taken it, he continued, to familiarize them with all the situations and to give them some idea where they might stand on each. Now he wanted the group to discuss each question in turn and arrive at a unanimous decision on each. This time they could not return to a question, but rather had to discuss each one until the group decision was reached before going on to the next. When the group reached its decision on a question, all subjects were to mark it on their questionnaires in order to have a record. The group would be completely on its own, the experimenter not participating in the discussion at all.

The experimenter then retired to the other end of the table in order to be as far from the group as possible. A question that often arose before discussion had started was what to do if a deadlock occurs. The experimenter's standard reply was:

> Most groups are able to come to some decision if those who disagree will restate their reasons, and if the problem is reread carefully.

Most groups succeeded in reaching a unanimous decision on most items, although an occasional deadlock did occur on one or another item. The group discussions were of such a nature as to indicate that the participants were highly involved in the decision tasks.

Postdiscussion individual decisions. After the discussion was over, the experimenter proceeded to ask the group members to spread apart for some further individual work and to take their questionnaires and identification placards with them. In standard instructions, he requested them to go back over the situations and indicate their own present personal decisions with a "P." He noted that while in some cases the subjects may have agreed with the group decision, in other cases they may have disagreed with it. In the former cases the P would be placed on the same line as the check mark; in the latter cases, on a different line.

While the consensual decisions by the group would indicate the public effect of the discussion process, the private postdiscussion decisions made

[3] Of the 14 male groups, 13 contained six subjects each, and one contained five subjects. A subject in one of the six-person male groups misunderstood instructions for the prediscussion individual decisions, so that his decision scores were removed prior to analysis. All 14 of the female groups contained six subjects each. A subject in each of 2 female groups misunderstood instructions for the prediscussion individual decisions, so that the decision scores of these two females were removed prior to analysis.

once again on an individual basis would indicate whether the discussion process had influenced covert acceptance as well as public compliance.

Rankings for influence and popularity. After the postdiscussion individual decisions had been made, a ranking sheet was passed out to each subject requesting that he rank everyone in the group (identified by their letter placards), including himself, in terms of how much each influenced the final group decision. Then each subject was requested to rank everyone in the group (except, of course, himself) in terms of how much he would like to become better acquainted with each.

The rankings for influence provided the information needed for examining possible relationships between strength of individual risk taking or conservatism tendencies, on the one hand, and degree of influence in the group, on the other. If such relationships existed, it seemed to be of interest to determine whether they were specific to perceived influence or would prove to be dependent upon the subject's popularity; hence the second set of rankings.

Secrecy instructions. After the ranking sheets were collected, the experimenter told the group that the research would be carried out in coming weeks, and that they could now appreciate why it would be important for the content of the experiment to be kept secret, since a person who even knew that the group would be discussing the same questions which he had filled out individually would have a tendency to mark logically defensible answers instead of his true opinion, etc. The subjects therefore all were sworn to secrecy. Various indications suggest that this pledge was faithfully kept.

Post-postdiscussion individual decisions. A further session of individual decision making took place approximately 2–6 weeks later for some subjects. These subjects individually were given the conservatism-risk questionnaire a third time and were asked to reconsider the situations. The standard instructions emphasized that the experimenter was not interested in testing the subject's memory, but rather wanted the subject truly to *reconsider* each situation. The instructions thus oriented the subjects away from simply trying to recall their prior decisions. Each subject was paid for this further work.

Control Condition

Subjects. Control subjects were obtained in the same way as the experimental subjects, and likewise received remuneration for their work. The controls were signed up to participate in two sessions: the first to last about 20 minutes; the second, exactly 1 week later, to last about 15 minutes. A total of 51 subjects participated in the control condition—24 males and 27 females. Like the experimental subjects, the controls were liberal arts students enrolled in summer session courses at the University of Colorado in Boulder.

First individual decision session. The first session was identical to the prediscussion individual decision part of the experimental condition. From six to eight subjects of the same sex, scheduled for the same time, filled out the conservatism-risk instrument while sitting together in physical conditions identical to those of the experimental subjects and at approximately the same time of day as the experimental subjects had worked. Exactly the same instructions were provided as had been given the experimental subjects.

After the first session, the control subjects were sworn to secrecy. They also were told that they would be taking a similar questionnaire the next week, and that it was extremely important that they not discuss it with one another nor with anyone else, since such discussion might affect the way they filled out next week's questionnaire.

Second individual decision session. The same control subjects who had participated in a particular first individual decision session came back exactly 1 week later. After checking that no discussion had taken place in the intervening week among the controls, the experimenter handed out new copies of the questionnaire and explained that this questionnaire was identical to the one taken last week. Each subject was requested to go back over the situations and reconsider them, the experimenter emphasizing that he was not interested in testing the subject's memory but rather wanted the subject truly to *reconsider* each situation. The instructions were so designed, therefore, as to dissuade the subject from assuming that the most socially acceptable thing to do would be to try to make the same decisions that he had made a week ago. Change was encouraged rather than discouraged. Control subjects were sworn to secrecy again at the end of the second session.

RESULTS

Consensual Group Decisions Compared with Prediscussion Individual Decisions

Tables 1 and 2 examine, for male and female groups, respectively, the significance of the conservatism difference between the mean of the prediscussion individual decisions made by the members of each group and that group's consensual decisions. The basic test is carried out using the total conservatism score, which consists of all 12 item scores combined. Tests also are carried out for each item separately.

In the case of the total score, a group's difference score is the sum of the 12 unanimous group decision scores minus the average of the prediscussion total individual decision scores for the six members.[4] Since larger scores indicate greater conservatism, a negative difference (or score decrease) indicates a

[4] Any deadlocked item is, of course, not included in either term for the group in question.

TABLE 1

SIGNIFICANCE OF CONSERVATISM DIFFERENCE BETWEEN MEAN OF PREDISCUSSION INDIVIDUAL DECISIONS FOR A GROUP'S MEMBERS AND GROUP'S CONSENSUAL DECISION: MALES

Item	Mean difference[a]	Number of groups[b]	t
All combined	−9.4	14	6.46****
1	−1.0	14	4.34****
2	−0.2	14	<1.00
3	−1.1	13	2.19*
4	−1.8	13	6.18****
5	+0.1	13	<1.00
6	−1.2	13	3.35**
7	−2.0	14	9.64****
8	−1.1	14	1.97
9	−1.0	10	3.67**
10	−0.4	13	<1.00
11	−1.1	12	4.37***
12	+0.8	11	2.34*

[a] In Tables 1, 2, 3, 4, 6, and 7, a negative difference signifies a risky shift, a positive difference signifies a conservative shift.

[b] In Tables 1 and 2, number of groups for an item is less than 14 when one or more groups deadlocked on that item. Any deadlocked item is, of course, not included when calculating scores for all items combined.

* $p < .05$.
** $p < .01$.
*** $p < .005$.
**** $p < .001$.

shift in the risky direction. A t test is used to determine whether the 14 difference scores for the groups of each sex are significantly different from zero (McNemar, 1955, pp. 108–109).[5] These total score data indicate a move in the risky direction significant beyond the .001 level for the 14 male groups, and a move in the risky direction significant beyond the .005 level for the 14 female groups. Furthermore, the degree of shift is not significantly different for the two sexes.

In the case of the scores for a single item, a group's difference score consists of the unanimous group decision on that item minus the average of the prediscussion individual decision scores on that item for the six members. Once again a negative difference or score decrease indicates a shift in the risky direction, and a t test is applied to determine whether the difference scores for all groups that reached a unanimous decision on the item in question are significantly different from zero. For both the male and female groups, we find that 10 of the 12 items show shifts in the risky direction, 7 of them significant in each case. Five of those 7 are the same for both sexes. Only 2 items show any indication for either sex of not sharing in the general shift toward greater risk taking: Items 5 and 12. It should be noted that these two items exhibited, in our previous research, the lowest correlations with the overall risk-conservatism score, suggesting that they are relatively impure measures of the psychological dimension being tapped by the other 10 items.

In sum, the evidence from Tables 1 and 2 indicates a strong move toward greater risk taking when groups arrive at unanimous decisions, compared with the risk levels ventured by the same persons in prediscussion individual decisions. Furthermore, this move toward greater risk taking obtains for females as well as for males.

A further question concerns the extent to which the risky shift is consistent from one group to another. Consider one example of several consistency tests that have been conducted, all of which yield highly similar results. Suppose we define a group as showing a risky shift from prediscussion individual decisions to consensual group decisions if the difference score for its total score, as defined above, is a negative one. Fourteen out of 14 male groups and 12 out of 14 female

TABLE 2

SIGNIFICANCE OF CONSERVATISM DIFFERENCE BETWEEN MEAN OF PREDISCUSSION INDIVIDUAL DECISIONS FOR A GROUP'S MEMBERS AND GROUP'S CONSENSUAL DECISION: FEMALES

Item	Mean difference	Number of groups	t
All combined	−9.4	14	3.91***
1	−1.0	13	4.17***
2	−0.6	14	1.65
3	−0.4	14	1.12
4	−1.4	14	2.60****
5	+0.7	14	1.90
6	−0.8	13	2.63****
7	−2.0	12	3.21**
8	−1.7	14	5.26*****
9	−0.8	12	1.19
10	−1.5	13	3.18**
11	−0.9	13	2.28*
12	+0.6	6	2.00

* $p < .05$.
** $p < .01$.
*** $p < .005$.
**** $p < .025$.
***** $p < .001$.

[5] All significance levels cited in this study are based on two-tailed tests.

TABLE 3

SIGNIFICANCE OF CONSERVATISM DIFFERENCE BETWEEN MEAN OF PREDISCUSSION INDIVIDUAL DECISIONS FOR A GROUP'S MEMBERS AND MEAN OF POSTDISCUSSION INDIVIDUAL DECISIONS FOR A GROUP'S MEMBERS: MALES

Item	Mean difference	Number of groups	t
All combined	−10.4	14	9.12****
1	− 1.0	14	4.32****
2	− 0.6	14	2.87*
3	− 1.1	14	3.04**
4	− 1.7	14	8.14****
5	+ 0.1	14	<1.00
6	− 1.1	14	3.79***
7	− 1.8	14	7.80****
8	− 1.1	14	3.54***
9	− 1.1	14	3.99***
10	− 0.3	14	<1.00
11	− 0.8	14	4.36****
12	+ 0.1	14	<1.00

* $p < .02$.
** $p < .01$.
*** $p < .005$.
**** $p < .001$.

groups are found to move in the risky direction, both results being very significant by a sign test. Such a finding demonstrates, therefore, that the risky shift phenomenon is quite consistent across groups.

Postdiscussion Individual Decisions Compared with Prediscussion Individual Decisions

In Tables 3 and 4 we present, once again for male and female groups, respectively, the significance of the difference between the mean of the prediscussion individual decisions and the mean of the postdiscussion individual decisions made by the members of each group. The basic test once again is provided by the total conservatism score, but tests also are presented for each item separately.

For the total score, a group's difference score consists of the average of the postdiscussion total individual decision scores for the members minus the average of the prediscussion total individual decision scores for the same members. Negative difference scores again indicate risky shifts, and a t test is applied to determine whether the 14 difference scores for the groups of each sex are significantly different from zero. We find, once again, a shift in the risky direction significant beyond the .001 level for the 14 male groups, and a risky shift significant beyond the .005 level for the 14 female groups. As before, the degree of shift is not significantly different for the two sexes.

Turning to the scores for each separate item, a group's difference score consists of the average of the postdiscussion individual decision scores on that item minus the average of the prediscussion individual decision scores on that item. With a negative difference score indicating a risky shift and a t test applied to indicate whether the 14 difference scores for each sex on an item are significantly different from zero, we find that 9 of the 12 items show separate significant shifts in the risky direction for the male groups (with one additional item shifting nonsignificantly in the same direction), and that 8 of the 12 items show separate significant shifts toward greater risk taking for the female groups (with two additional items shifting nonsignificantly in that direction). The 8 items showing significant risky shifts for the females are among the 9 showing significant risky shifts for the males. Items 5 and 12 once again are the only ones for either sex showing any indication of not sharing in the general shift toward greater risk taking found in both sexes.

There is clear evidence, therefore, that

TABLE 4

SIGNIFICANCE OF CONSERVATISM DIFFERENCE BETWEEN MEAN OF PREDISCUSSION INDIVIDUAL DECISIONS FOR A GROUP'S MEMBERS AND MEAN OF POSTDISCUSSION INDIVIDUAL DECISIONS FOR A GROUP'S MEMBERS: FEMALES

Item	Mean difference	Number of groups	t
All combined	−8.2	14	3.67**
1	−0.9	14	5.09****
2	−0.7	14	2.67*
3	−0.6	14	2.58***
4	−1.4	14	3.40**
5	+0.6	14	1.85
6	−0.8	14	2.90*
7	−1.7	14	3.56**
8	−1.2	14	4.44****
9	−0.5	14	<1.00
10	−0.7	14	1.95
11	−0.9	14	2.89*
12	+0.7	14	3.66**

* $p < .02$.
** $p < .005$.
*** $p < .025$.
**** $p < .001$.

TABLE 5

COMPARABILITY OF EXPERIMENTAL AND CONTROL SUBJECTS IN INITIAL CONSERVATISM AND AGE

Subject	Males		Females	
	M	*N*	*M*	*N*
	Mean Initial Overall Conservatism			
Experimental	66.9	82[a]	65.6	82[a]
Control	68.3	24	64.6	27
t	0.41		0.34	
	Mean Age			
Experimental	20.7	82[b]	20.3	84
Control	21.0	24	20.7	27
t	0.41		0.67	

[a] Initial overall conservatism scores were available for 164 of the experimental subjects. See Footnote 2 in text.
[b] One subject forgot to list his age, and one group contained five rather than six subjects.

postdiscussion individual decisions exhibit a strong move toward greater risk taking when compared with prediscussion individual decisions arrived at by the same persons, and do so for both sexes. The group discussion process, in other words, seems to have an effect on private attitudes (postdiscussion individual decisions) that is just as significant as its effect on publicly expressed views (unanimous group decisions).

Once again we may inquire about the extent to which the risky shift is consistent from group to group. Several consistency tests have been carried out, all yielding highly similar results. As an example, suppose we define a group as exhibiting a shift in the risky direction from prediscussion to postdiscussion individual decisions if the difference score for its total score, as defined in this section, is a negative one. Fourteen out of 14 male groups and 12 out of 14 female groups are found to shift in the risky direction, both results being quite significant by a sign test. Such a finding demonstrates, therefore, that the risky shift pheomenon is quite consistent across groups in regard to covert acceptance as well as overt compliance.

Control Subjects

To insure that the move toward greater risk taking just described actually is a result of the group discussion process, we must turn to the findings for the control subjects. The comparability of control and experimental subjects is indicated in Table 5. We note that, in the case both of males and females, the experimental and control subjects have approximately the same initial total conservatism scores, and also are approximately the same in age.[6] Item-by-item comparisons of experimental and control subjects of each sex on initial conservatism scores also were carried out and show that controls and experimentals within sex obtain highly similar scores.

In Tables 6 and 7 we present, for male and female control subjects, respectively, the significance of the difference between decisions made during the first and the second sessions. It will be recalled that one week intervened between these two sessions, and that instructions for the second session requested the subjects not to try simply to remember what they had marked before, but to reconsider their decisions. It is evident that the total conservatism score shows no shift from first to second session for either sex. Turning to the separate tests carried out on each item, we find that none of the 12 items shows a significant shift for the males, and only 1 of the 12 items shows a significant shift for the females. When no group discussion and

[6] It might also be mentioned that, in confirmation of earlier findings (Wallach & Kogan, 1959, 1961), there is no sex difference in initial total conservatism scores for either the experimental or the control subjects.

TABLE 6

SIGNIFICANCE OF CONSERVATISM DIFFERENCE BETWEEN FIRST AND SECOND DECISIONS BY MALE CONTROL SUBJECTS

Item	Mean difference	Number of subjects	*t*[a]
All combined	+1.5	24	<1.00
1	+0.4	24	<1.00
2	−0.3	24	<1.00
3	+0.3	24	<1.00
4	+0.8	24	2.00
5	−0.4	24	1.06
6	0.0	24	<1.00
7	+0.4	24	1.03
8	+0.5	24	1.63
9	−0.1	24	<1.00
10	+0.1	24	<1.00
11	+0.1	24	<1.00
12	−0.4	24	1.42

[a] All *t* values *ns*.

TABLE 7

SIGNIFICANCE OF CONSERVATISM DIFFERENCE BETWEEN FIRST AND SECOND DECISIONS BY FEMALE CONTROL SUBJECTS

Item	Mean difference	Number of subjects	t
All combined	−2.2	27	1.26
1	−0.4	27	<1.00
2	−0.2	27	<1.00
3	−1.0	27	2.61*
4	−0.4	27	1.12
5	−0.3	27	<1.00
6	−0.2	27	<1.00
7	0.0	27	<1.00
8	0.0	27	<1.00
9	+0.2	27	<1.00
10	+0.3	27	1.03
11	−0.3	27	<1.00
12	+0.1	27	<1.00

* $p < .02$.

achievement of group consensus intervenes, then, there is no systematic shift toward to the findings for the control subjects. The greater risk taking or greater conservatism, and this despite instructions that encourage shifts by emphasizing that we are not interested in the subjects' memories.

The data for the control subjects also provide us with an opportunity for determining the test-retest reliability of the conservatism-risk instrument, with one week intervening and under instructions that encourage change rather than constancy. For the 24 male subjects, the product-moment correlation coefficient between total conservatism scores in the first and second sessions is .78. For the 27 female subjects, the same correlation coefficient is .82. Test-retest reliability of the instrument, therefore, is quite high.

Prediscussion Risk Taking and Influence in the Group

Our data concerning perceived influence within the group consisted in each individual's ranking of all group members, including himself, in terms of how much each influenced the group's decisions. A first question to ask of these influence rankings is: How consistent are they from member to member within a group? To determine the degree of agreement among a group's members in their rankings of one another for influence, Kendall's coefficient of concordance (Siegel, 1956, pp. 229–238) was applied to each group's influence rankings. If the members of a group agree

TABLE 8

DEGREE OF AGREEMENT AMONG GROUP MEMBERS IN RANKINGS OF ONE ANOTHER FOR INFLUENCE[a]

Group	Males		Group	Females	
	N	W		N	W
1	6	.64**	1	6	.85**
2	6	.55**	2	6	.61**
3	6	.74**	3	6	.31
4	6	.72**	4	6	.79**
5	6	.70**	5	6	.47**
6	6	.50**	6	6	.67**
7	5	.56*	7	6	.13
8	6	.50**	8	6	.59**
9	6	.62**	9	6	.59**
10	6	.66**	10	6	.69**
11	6	.66**	11	6	.83**
12	6	.55**	12	6	.80**
13	6	.54**	13	6	.70**
14	6	.73**	14	6	.30

[a] Kendall's coefficient of concordance.
* $p < .05$.
** $p < .01$.

regarding who among themselves are more influential and who less so, then W will be significantly large. Table 8 presents the results of these tests for all 28 groups. It is evident that agreement in influence rankings is quite high: the degree of agreement is significant for all 14 of the male groups, and for 11 of the 14 female groups.

Given this high agreement among group members in their rankings of one another for influence, an approximate overall estimate

TABLE 9

PRODUCT-MOMENT CORRELATIONS AMONG INITIAL CONSERVATISM, INFLUENCE, AND POPULARITY[a]

	Males ($N = 82$)[b]	Females ($N = 82$)[b]
	r	r
Initial overall risk taking and influence	.32****	.22*
Initial overall risk taking and popularity	.15	−.04
Influence and popularity	.72*****	.54*****
Initial overall risk taking and influence, popularity held constant[c]	.30***	.28**

[a] Small score values signify greater risk taking, greater influence, and greater popularity.
[b] While all influence and popularity scores are based on the 167 subjects in the experimental condition, the correlations are based on the 164 of those subjects for whom initial overall risk taking scores were available.
[c] Partial correlation coefficients.
* $p < .05$.
** $p < .02$.
*** $p < .01$.
**** $p < .005$.
***** $p < .001$.

of degree of influence for a given group member was obtained by averaging the influence ranks that had been assigned to that person by all members of the group (including that person). The lower the average, the greater that subject's perceived influence (i.e., the higher the assigned influence ranks for that person). These average influence scores for the subjects of each sex were correlated with the initial total conservatism scores obtained by the same subjects. The resulting product-moment correlation coefficients are shown in Table 9. They are significant beyond the .005 and .05 levels for the 82 males and the 82 females, respectively: persons higher in initial risk taking are rated as having more influence on the group decisions.

Average popularity scores for each group member were constructed by averaging the popularity rankings assigned by all the other members of the group. We note in Table 9 that there emerges a very strong relationship between this average popularity score and the average influence score for both the male and the female group members: persons rated high in influence also tend to be rated high in popularity. This general relationship has, of course, been known for some time (see, e.g., Back, 1951; Horowitz, Lyons, & Perlmutter, 1951; Tagiuri & Kogan, 1960), so that our obtaining it here increases our confidence in the respective measures being used to assess influence and popularity. It is further evident in Table 9, however, that degree of initial risk taking is *not* related to degree of popularity within the group for either sex.

Finally, we also find from Table 9 that risk taking and influence are significantly related for each sex when popularity ratings are held constant. The partial correlation coefficients are significant beyond the .01 and .02 levels for the males and females, respectively. It is evident, therefore, that the relationships obtained for both sexes between degree of initial risk taking and degree of influence on group decisions are not dependent upon members' popularity.

Maintenance of the Risky Shift over a Subsequent Period of Time

An interesting further question concerns the extent to which the shift toward greater risk taking, which we have found to result from group discussion, is maintained over a subsequent period of time. We were able to gather evidence on this point for males but not for females. In the case of the former, but not in the case of the latter, a random sample of subjects from the original groups could be obtained for further study. The 22 males who were available for further work were approximately evenly distributed among the 14 original male groups. After a time interval of roughly 2–6 weeks had elapsed since the group session, these subjects individually were given the conservatism-risk questionnaire a third time, as described in the section on procedure.

The comparability of the random male subsample of 22 to the original male experimental condition sample of 82 is evident from the following data on total conservatism scores. The mean prediscussion total conservatism score was 66.9 for the sample of 82, and also was 66.9 for the subsample of 22. The mean postdiscussion total conservatism score, in turn, was 56.6 for the whole sample and 56.2 for the subsample. The t test of the difference scores had yielded a t significant beyond the .001 level ($t = 9.12$) for the whole sample, and it also yielded a t significant beyond the .001 level ($t = 4.70$) for the subsample.

Turning now to the total conservatism scores obtained by this subsample when they took the questionnaire again 2–6 weeks after the group discussion (call these scores the "post-postdiscussion" individual decisions), the mean score is 54.6. The mean of the difference scores obtained by subtracting each subject's prediscussion total conservatism score from his post-postdiscussion total conservatism score is -12.3, with a t test of these difference scores yielding a t value of 4.92 ($p < .001$), hence indicating a risky shift from the prediscussion individual decisions to the post-postdiscussion individual decisions. The mean of the difference scores obtained, in turn, by subtracting each subject's postdiscussion total conservatism score from his post-postdiscussion total conservatism score is only -1.6, and a t test of these difference scores is not significant, hence indicating no further change from the postdiscussion individual decisions to the post-

postdiscussion individual decisions. Item-by-item analyses tell the same story: the only significant item shifts are risky ones, and they are as strong from prediscussion to post-postdiscussion sessions as they are from prediscussion to postdiscussion sessions.

In sum, the data available on the point indicate that the shift in the risky direction found to occur as a result of the group discussion process is maintained over a subsequent period of time.

DISCUSSION AND CONCLUSIONS

The following conclusions may be drawn from the preceding evidence:

1. Unanimous group decisions concerning matters of risk show a shift toward greater risk taking when compared with prediscussion individual decisions made by the same persons and concerning the same matters. This holds for both sexes.

2. Postdiscussion individual decisions that follow unanimous group decisions exhibit the same kind of shift toward greater risk taking as appears in the group decisions. This is the case for both sexes. Covert acceptance as well as overt compliance, thus, are affected in the same manner by the discussion process.

3. This shift toward greater risk taking as a result of the discussion process is still maintained when 2–6 weeks have elapsed since the discussion occurred. Evidence on this point was available only for males.

4. No shift in risk taking level of individual decisions occurs over time in the absence of the discussion process. This holds for both sexes.

5. There is a positive relationship between degree of risk taking in prediscussion individual decisions and the extent to which group members are perceived by one another as influencing group decisions. This relationship is specific to judgments of influence, in that it obtains when judgments of popularity are held constant, and also no relationship is found between prediscussion individual risk taking and the extent to which group members are judged to be popular. These statements all hold for both sexes.

The present study indicates, then, that group interaction and achievement of consensus concerning decisions on matters of risk eventuate in a willingness to make decisions that are more risky than those that would be made in the absence of such interaction. Furthermore, although initial ascribed status levels of the group members are equal, it is found that persons with stronger individual risk taking proclivities tend to become more influential in the group than persons who are more conservative. Two alternative interpretations of these findings can be suggested; one more group centered, the other more person centered: It is possible that there is at work in these groups a process of diffusion or spreading of responsibility as a result of knowing that one's decisions are being made jointly with others rather than alone. Increased willingness to take risk would eventuate from this decreased feeling of personal responsibility. That initial risk taking and judged influence within the group are positively related could well occur as a consequence of this process, since one of its effects would be for the views of high risk takers to be given more weight by the rest of the group. Alternatively, the fact that high risk takers exert more influence may be a cause of the group's movement toward greater risk taking. It is possible that high risk takers are also more likely to take the initiative in social situations. Of course, these two interpretations are not necessarily mutually exclusive. Both of them may contribute to the group effect.

That females as well as males show the same change toward greater risk taking as a result of the group interaction condition, and that the samples of both sexes were liberal arts university students, renders it unlikely that the results can be explained on the basis of reinforcement by others of one's expectation as to whether one's appropriate role is to be more or less of a risk taker. We noted earlier that Stoner (1961) found a move toward greater risk taking in group as compared to individual decision making by male graduate students of industrial management, and we pointed out that this result might be accounted for in terms of the professional role that they had assigned themselves by becoming graduate students in a business school. Presence of peers might be expected to increase the salience of their business

manager role, and a greater willingness to take risks in decision making might well be perceived as one of the attributes of that role. Such a role expectation interpretation is ruled out for the present study, however, through our use of liberal arts students as subjects. In addition, the possibility of explaining the results in terms of males' perceiving their appropriate role as one of willingness to be bold and daring, and being reinforced in this view by interaction with other like-minded males, is ruled out by the present study's obtaining the same results for females as for males. This outcome would not be expected if the findings depended on sex linked role expectations as to whether one should be more risky or more conservative. This outcome also, of course, rules out interpretation in terms of any possible sex linked differences in major fields of study.

That the group induced move toward greater risk taking in individual decisions is still maintained 2–6 weeks after the discussion, provides evidence, incidentally, which supports Lewin's (1947) view that "group carried" attitudinal changes maintain themselves (see also Pelz, 1958).

REFERENCES

ATTHOWE, J. M., JR. Interpersonal decision making: The resolution of a dyadic conflict. *J. abnorm. soc. Psychol.*, 1961, **62**, 114–119.

BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

BROWN, R. W. Mass phenomena. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. *Special fields and applications*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 833–876.

CARTWRIGHT, D., & ZANDER, A. (Eds.) *Group dynamics*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.

FESTINGER, L. An analysis of compliant behavior. In M. Sherif & M. O. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953. Pp. 232–255.

HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57–76.

HUNT, E. B., & ROWE, R. R. Group and individual economic decision making in risk conditions. In D. W. Taylor (Ed.), *Experiments on decision making and other studies*. Arlington, Va.: Armed Services Technical Information Agency, 1960. Pp. 21–25. (Technical Report No. 6, AD 253952)

KELLEY, H. H., & THIBAUT, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. *Special fields and applications*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 735–785.

KOGAN, N., & WALLACH, M. A. The effect of anxiety on relations between subjective age and caution in an older sample. In P. H. Hoch & J. Zubin (Eds.), *Psychopathology of aging*. New York: Grune & Stratton, 1961. Pp. 123–135.

KOMAROVSKY, MIRRA. Functional analysis of sex roles. *Amer. sociol. Rev.*, 1950, **15**, 508–516.

LEWIN, K. Frontiers in group dynamics. *Hum. Relat.*, 1947, **1**, 2–38.

LONERGAN, B. G., & MCCLINTOCK, C. G. Effects of group membership on risk-taking behavior. *Psychol. Rep.*, 1961, **8**, 447–455.

MCNEMAR, Q. *Psychological statistics*. (Rev. ed.) New York: Wiley, 1955.

MILNER, ESTHER. Effects of sex role and social status on the early adolescent personality. *Genet. psychol. Monogr.*, 1949, **40**, 231–325.

OSBORN, A. F. *Applied imagination*. New York: Scribner, 1957.

PELZ, EDITH B. Some factors in "group decision." In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (3rd ed.) New York: Holt, 1958. Pp. 212–219.

SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

STONER, J. A. F. A comparison of individual and group decisions involving risk. Unpublished master's thesis, Massachusetts Institute of Technology, School of Industrial Management, 1961.

TAGIURI, R., & KOGAN, N. Personal preference and the attribution of influence in small groups. *J. Pers.*, 1960, **28**, 257–265.

TAYLOR, D. W., BERRY, P. C., & BLOCK, C. H. Does group participation when using brainstorming facilitate or inhibit creative thinking? *Admin. sci. Quart.*, 1958, **3**, 23–47.

THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups*. New York: Wiley, 1959.

TURNER, R. H., & KILLIAN, L. M. (Eds.) *Collective behavior*. Englewood Cliffs, N. J.: Prentice-Hall, 1957.

WALLACH, M. A., & CARON, A. J. Attribute criteriality and sex-linked conservatism as determinants of psychological similarity. *J. abnorm. soc. Psychol.*, 1959, **59**, 43–50.

WALLACH, M. A., & KOGAN, N. Sex differences and judgment processes. *J. Pers.*, 1959, **27**, 555–564.

WALLACH, M .A., & KOGAN, N. Aspects of judgment and decision making: Interrelationships and changes with age. *Behav. Sci.*, 1961, **6**, 23–36.

WHYTE, W. H., JR. *The organization man*. New York: Simon & Schuster, 1956.

(Received October 17, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 87–94

FREE ASSOCIATION PATTERNS AS A FACTOR RELATING TO SEMANTIC DIFFERENTIAL RESPONSES

ROBERT F. TERWILLIGER

Douglass College

Ross and Levy (1960) have demonstrated that it is possible to describe the distribution of free associations to a stimulus consisting of a single adjective by means of the information or H statistic. This statistic is defined as follows (Attneave, 1959):

$$H = -\Sigma p \log_2 p$$

H stands for the amount of information in "bits," p is the probability of a response, and the summation is made over all possible responses to the stimulus in question. The characteristics of this descriptive statistic are such that it will increase under two conditions. An increase in H will occur as the number of different responses increases. And an increase in H will occur as the probabilities of the various responses to a given stimulus become more equal.

To arrive at an H value descriptive of the free associations to a given stimulus word, Ross and Levy (1960) required each number of a large sample of subjects to give one free association response to each stimulus word. In contrast to the usual free association procedures, the responses required of the subjects were not words but were patterns produced by arranging colored cards in a 3×3 matrix form. For a given stimulus word, the probability of occurrence of each of the various different free association responses was calculated by dividing the number of times that response occurred by the total number of responses. The probabilities for each different free association for a single stimulus word were entered into the equation given above and summed, the sum being the H value for that stimulus word.

Ross and Levy (1960) state that their method measures the information or uncertainty of the stimulus words. This may indeed be the case, but it is also a confusing and complicated issue. For the purposes of this paper, it is only necessary to consider the empirical matter. The H statistic can describe the distribution of free associations given to a stimulus word. If two words have different H values, then they have different distributions of free associations. The words "information" and "uncertainty" will not be used in this paper. Only the term H will be used and only as a descriptive statistic with no necessary theoretical connotations.

Ross and Levy (1960) used their method to criticize the concept of semantic space proposed by Osgood, Suci, and Tannenbaum (1957). To the latter theorists, the meaning of a word may be described by assigning it to a point in a three dimensional space. The dimensions of this space are three aspects of meaning: evaluation, potency, and activity. Each of these three orthogonal dimensions is bipolar. At one end of a dimension there is one particular extreme meaning. The dimension then runs through a neutral point and on to an equally extreme but completely opposite meaning.

If the Semantic Differential is to be valid in terms of accurately assigning a word to a place in the semantic space, it is necessary that the opposite adjectives used in the scales define equal amounts of the semantic space. Ross and Levy (1960) have shown in their study that the opposite adjectives used in the Semantic Differential have different H values. From this they argue that these adjectives do not define equal amounts of semantic space and therefore that the assumptions underlying the Semantic Differential are erroneous. Ross and Levy, however, have not shown that the H value of a word makes any difference in terms of its Semantic Differential rating. Only if one can show that Semantic Differential responses vary as a function of H can one use H values as critical of the Semantic Differential's ability to assign words accurately to semantic space. It will be the purpose of this study to show that the H value of a word

will influence the rating it receives on the Semantic Differential.

The first hypothesis to be tested is that a word with a high *H* value will be given a more meaningful rating on the Semantic Differential than will a word with a low *H* value. A "meaningful rating on the Semantic Differential" refers to that aspect of such ratings which Osgood et al. (1957) call *polarity*. The polarity of a rating is the mathematically absolute rating or the distance of the rating from the neutral point on a scale. Since the Semantic Differential scales run from ratings of +3 through neutral (0) to −3, the higher the rating, the higher the polarity of the word. The sign of the rating is ignored. It is predicted, then, that high *H* value words should have greater polarity on the Semantic Differential than low *H* value words.[1] This prediction should hold equally well for other types of verbal stimuli, e.g., phrases consisting of a noun modified by an adjective. It is possible to determine *H* values for such phrases and also possible to rate them on the Semantic Differential. Both the single word and phrase aspects of this hypothesis will be tested.

Osgood et al. (1957) have presented a hypothesis concerning the changes in meaning which take place through the grammatical operation of modification. An adjective alone has a particular meaning, as does a noun alone. If, however, the noun is modified by the adjective, the meaning of the phrase is different from either of its constituent words alone. Osgood et al. refer to the hypothesis which predicts meaning changes due to modification as the congruity hypothesis. This hypothesis assumes that the meaning of a phrase as measured by the Semantic Differential will be a weighted average of the meanings of its constituent words. The formula[2] for predicting the Semantic Differential rating of an adjective-noun phrase from the ratings of the adjective and noun alone is given below. Such predictions are made separately for each subject and each scale of the Semantic Differential.

$$\text{Predicted rating} = \frac{|p_a|}{|p_a| + |p_n|} p_a + \frac{|p_n|}{|p_a| + |p_n|} p_n$$

Here $|p_n|$ stands for the polarity of the noun on a given scale of the Semantic Differential, $|p_a|$ is the polarity of the adjective on the same scale, while p_a and p_n are the ratings of the adjective and noun including the sign of the rating.

It is assumed that the *H* values of the words in a phrase will relate to the accuracy of the congruity hypothesis predictions. It is possible to calculate the discrepancy between the rating of an adjective-noun phrase as predicted from the congruity hypothesis and the actual rating given that phrase by the subject on the scale concerned. The direction or sign of this discrepancy will be ignored. It is predicted that when a noun and an adjective used in a phrase differ highly in their respective *H* values, the discrepancy between the predicted congruity hypothesis rating and the obtained rating of the phrase will be large. When such differences in *H* values are small, the discrepancy between the predicted and obtained ratings of that phrase will be small.

There is an alternative way of testing the same hypothesis on the same set of data. Any predicted congruity rating may be categorized as being accurate or inaccurate in view of the obtained ratings of the phrase. Among phrases whose member words differ greatly in their *H* values, it is predicted that the congruity hypothesis will make fewer accurate predictions than among phrases whose words

[1] The terms "high" and "low" or their equivalents are used throughout the discussion of the hypotheses. They may be defined operationally as follows: given a group of words measured on a particular dimension, e.g., *H* value, high refers to words above the median on that dimension and low refers to words below the median.

[2] A brief consideration of this equation will demonstrate that it is the more polar word which has greater weight in determining what the rating of the phrase should be. If words are equal and opposite in polarity, they will cancel each other out, the predicted rating being zero or neutral. If one word is neutral and the other has some polarity, the rating of the phrase will be that of the polar word. In between these extremes it can be seen that the more polar word is given greater weight. As a single example, a phrase consisting of a noun with a rating of +3 modified by an adjective with a rating of −2 has a predicted rating of +1. Straight averaging would produce a predicted rating of +.5.

differ less in their *H* values. Both of the alternative tests of the hypothesis will be made. The difference between the two tests consists simply in treating the data continuously in the first case and discontinuously in the second.

Any discrepancy from the predicted congruity rating of a phrase may be said to have a direction. That is, the obtained rating will diverge from the predicted in the direction of the rating of one or the other of the individual words in the phrase. The general hypothesis to be tested is that discrepancies from the predicted congruity ratings will be in the direction of the rating of the word in the phrase with the lower *H* value, and that such discrepancies will increase as the difference in *H* values of the words in the phrase increases. Strictly speaking, it is necessary to conceive of a distribution of discrepancies from predicted congruity ratings. Chance discrepancies from the predicted rating will always exist. The mean of these chance discrepancies will be zero when there is no difference in the *H* values of the words in the phrase. As the words in a phrase become different in their *H* values, this mean discrepancy will shift more and more toward the rating of the word in the phrase with the lower *H* value. This means that more discrepancies will be toward the rating of the low *H* value word in the phrase and fewer will be toward the rating of the higher *H* value word. Given two sets of phrases, one whose member words differ slightly in their *H* values and one whose member words differ greatly in their *H* values, the following relationships should hold:

1. The amount of discrepancy toward the rating of the low *H* value word should be greater in those phrases whose member words differ more in their *H* values than in phrases whose member words differ less in their *H* values.

2. The amount of discrepancy toward the rating of the higher *H* value word should be greater in those phrases whose member words differ less in their *H* values than in phrases whose member words differ more in their *H* values.

3. The net discrepancy from congruity is defined as the total discrepancy toward the rating of the low *H* value word in the phrase minus the total discrepancy toward the rating of the high *H* value word. Among phrases whose member words differ greatly in their *H* values, there will be greater net discrepancy from the predicted congruity ratings toward the rating of the low *H* value word in the phrase than among phrases whose words differ less in their *H* values.

Note that three different tests of the basic hypothesis have been proposed. These are not independent tests or experimental replications. They are simply alternative ways of analyzing the same set of data. The data gathered in this study will be analyzed in all three ways.

METHOD

Measuring H *Values*

The *H* values of the words used were assessed in a manner analogous to that used by Ross and Levy (1960). The subjects were given booklets, on each page of which was a grid of nine squares arranged in three rows and three columns. To the left of this grid was a word, either a noun or an adjective. The subjects were given the following instructions:

> Each of the sheets on which you are to work has on it a square which has been divided up into nine small squares. These small squares are arranged in three rows of three squares each. You will be asked to make patterns by filling in some, none, or all of these small squares. You are free to make the patterns exactly as you wish. You may fill in any number of the squares in any order you wish. You may leave them all blank or fill them all in if you think this is more appropriate. If you choose to fill in one of these small squares, you must fill in the whole thing—you cannot subdivide it into smaller units than those already there.
>
> To the left of each of these matrices of squares you will see a word, either an adjective or a noun. Your job is to draw a pattern by filling in squares (or not filling them in, as you choose) such that this pattern, in your opinion, represents the word to the left of it as closely as the situation allows. For example, if the word were REGULAR, you are to draw a pattern which you would call "regular." You may decide that none of the possible patterns is perfectly regular, in which case you should draw the one that comes closest to it. If the word were REGULARITY, you should draw the pattern which comes as close as possible to typifying "regularity."

Twelve words in all were used, six adjectives and six nouns. The adjectives were polar opposites, e.g., BEAUTIFUL and UGLY, and the nouns were the complimentary adjectival nouns, e.g., BEAUTY and UGLINESS. The test was administered during a class period to three sections of introductory psychology. A total of 108 female undergraduates were tested.

TABLE 1

THE *H* VALUES OF STIMULUS WORDS

Stimulus word	*H*
Adjective	
BEAUTIFUL	2.6262
UGLY	4.3840
COMMON	2.8111
RARE	4.1195
SIMPLE	2.8609
COMPLEX	4.4845
Noun	
BEAUTY	3.3086
UGLINESS	4.1885
COMMONNESS	2.9312
RARITY	4.4018
SIMPLENESS	2.7451
COMPLEXITY	4.1134

It was necessary to establish a criterion for evaluating which of the patterns produced by the subjects constituted similar patterns and which patterns were different. Following Ross and Levy (1960), the following rules of combination were used: black-white distinctions were disregarded regarding whether filled-in or blank squares were in the majority. Thus, for example, a pattern in which all nine squares were filled in was treated as identical to one in which all nine squares were left blank. All 90° rotations of a given pattern, all mirror images of the pattern, and all 90° rotations of mirror images of the pattern were treated as identical to the pattern in question. Using the above rules for evaluating what were identical patterns, it was found that there remained a total of 51 possible different patterns which could be produced. For each word, the proportion of subjects who produced each of the possible different patterns was determined. These proportions were entered into the equation for the *H* statistics as given by Attneave (1959) and summed. This gave an *H* value for each word. The *H* values of each of the 12 words used in this study are given in Table 1.

The *H* values of the adjective-noun phrases were determined in a somewhat arbitrary manner. It was assumed that a reasonable indicator of the *H* value of a phrase would be given by the sum of the *H* values of its member words. Since it is possible for two stimuli to "share information" or *H* value, such straight addition may yield an overestimate or at least an inaccurate estimate of the true *H* value of the phrase. However, the design of the study did not permit the more sophisticated analysis of accounting for shared *H*. The difference in *H* value of the two words in a phrase was determined by subtracting the two individual *H* values. The above criticism holds for this measure as well.

Semantic Differential

Fifteen volunteer subjects (female) were given a form of the Semantic Differential in which each word was to be rated on 12 different scales. Table 2 indicates the polar adjectives used on this form of the Semantic Differential. Each subject was required to attend two one-hour sessions held one week apart. In the first of these sessions the subjects were required to rate the 12 individual adjectives and nouns. They were given booklets on each page of which were the 12 Semantic Differential scales. At the top of each page was the word to be rated. Two practice words were given first to acquaint them with the use of the Semantic Differential. In the second session, similar booklets were given to the subject. Here the stimuli to be rated were adjective-noun phrases. Thirty possible phrases could be prepared from the adjectives and nouns by pairing each adjective with every noun except that from which it was derived. Thus, for example, BEAUTIFUL was paired with all nouns except BEAUTY. Some of these adjective-noun combinations are quite unusual, e.g., BEAUTIFUL UGLINESS. These odd phrases were necessitated by the pictorial method of assessing the *H* value of the words. It was necessary to use words which could reasonably be expected to describe patterns, hence some odd phrase combinations were unavoidable.

From the Semantic Differential ratings of the individual words, the total polarity of each word was obtained by summing the absolute rating of each word over all subjects and all scales of the Semantic Differential. Predicted congruity ratings of the phrases were determined by entering the ratings of the individual words into the appropriate equation (Osgood et al., 1957). This was done separately for each subject and each scale of the Semantic Differential.

The total polarity of each phrase on the Semantic Differential was obtained by summing the absolute rating of the phrase over all subjects and all scales. Discrepancies from the predicted congruity ratings were obtained by subtracting the predicted from the obtained rating of the phrase separately for each subject and each scale. The sign of this difference was ignored. The total discrepancy from the predicted congruity ratings for a phrase was determined by summing these discrepancies over all subjects and all scales. For each subject and scale separately, it was noted whether the predicted congruity

TABLE 2

ADJECTIVE PAIRS USED IN SEMANTIC DIFFERENTIAL

Adjective pairs
Pleasant–Unpleasant
Active–Passive
Tight–Loose
Sociable–Unsociable
Bad–Good
Calm–Excitable
Soft–Hard
Clear–Hazy
Agreeable–Disagreeable
Emotional–Unemotional
Closed–Open
Intentional–Unintentional

TABLE 3

RESULTS[a] OF THE HYPOTHESES

Hypothesis		High	Low	Median	U
1. Polarity on Semantic Differential					
a. H value of words	High	4	2	312.5	7*
	Low	2	4	254.5	
b. H value of phrases	High	10	5	292	72.5
	Low	5	10	232	
2. Total discrepancy from predicted congruity rating					
a. Differences of H values of words in phrase	High	11	4	231	58**
	Low	4	11	216	
Total number of accurate congruity predictions					
b. Differences of H values of words in phrase	High	4	11	54	49**
	Low	11	4	71	
3. Total discrepancy from predicted congruity rating toward low H value word					
a. Differences of H values of words in phrase	High	10	5	112	51**
	Low	5	10	82	
Total discrepancy from predicted congruity rating toward high H value word					
b. Differences of H values of words in phrase	High	7	8	81	103.5
	Low	8	7	89	
Total net discrepancy from predicted congruity ratings toward low H value word					
Differences of H values of words in phrase	High	10	5	25	69.5*
	Low	5	10	−15	

[a] High and Low refer to above and below the median on the variable in question.
* $p \leq .05$.
** $p < .025$.

rating was accurate (1) or not (0). The total number of accurate predictions was determined for each phrase by summing over scales and subjects. In some cases, the congruity equation predicts a fractional rating. These could not be obtained in actual ratings, which were always whole numbers. In such cases, either of the two nearest whole number ratings was considered an accurate rating. For example, given a predicted rating of +2.5, an obtained rating of +2 or +3 would be considered accurate. This procedure was also used in determining the amount of discrepance by congruity. The predicted rating was always rounded to the number which would make the discrepancy smallest.

Each discrepancy from the congruity predictions was coded to indicate whether it was in the direction of the rating of the lower H value individual word. The total discrepancy toward the lower H value word was determined for a phrase by summing all discrepancies coded in this direction over all subjects and scales. In a similar manner, the total discrepancy toward the higher H value word was determined. The net discrepancy from congruity toward the lower H value word was determined by subtracting the total discrepancy toward the higher H value word from the total discrepancy toward the lower H value word. There were some discrepancies from predicted congruity ratings which could not be coded as to direction.[3] Such discrepancies were ignored for this analysis. They were, however, included in the determination of the total absolute discrepancy from congruity.

The data did not appear to meet the requirements for a parametric statistical analysis, nor was the assumption of linearity necessary for correlational analysis met; therefore, all hypotheses were tested with a nonparametric statistical device. The Mann-Whitney U test was used throughout.

RESULTS AND DISCUSSION

Hypothesis 1 states that high H value words and phrases will have greater total

[3] As an example of this, take an adjective with a rating of −2 and a noun with a rating of −2. The congruity hypothesis predicts a rating for the phrase of −2. If the obtained rating were something else, e.g., −3, the discrepancy could not be said to be in the direction of one of the words rather than the other. It was impossible to assign a direction to such a discrepancy.

polarity than will low *H* value words and phrases. Words and phrases were divided into high and low *H* value groups by splitting at the median *H* value. Table 3 indicates that in the case of the individual words, the hypothesis was supported at the .05 level of significance (one-tailed) while in the case of the phrases, the hypothesis was met at slightly greater than the .05 level. The obtained *U* was 72.5, while that necessary for significance was 72.0. Considering that two sets of ties existed in the data, thus adversely affecting the *U* value, it is reasonable to assume that this hypothesis stands supported.

Hypothesis 2 states that there will be greater total discrepancies from predicted congruity ratings among phrases in which the words are highly different in *H* value than among phrases in which the words are more similar in *H* value. Once again, phrases were divided into two groups, one in which the words were highly different in *H* value and one in which the words were less different in *H* value, by dividing at the median of differences in *H* value. Statistical analysis indicates that the hypothesis was supported at less than the .025 level of significance (see Table 3). The alternative test of this hypothesis states that among phrases which have high differences in the *H* values of their member words, there will be fewer accurate congruity predictions than among phrases which are low in such differences in *H* value. Once again the hypothesis was verified at less than the .025 level of significance.

Hypothesis 3 states that discrepancies from the predicted congruity ratings should be in the direction of the rating of the word in the phrase with the lower *H* value. There are three possible tests of this hypothesis. The first of these states that among phrases whose words differ greatly in *H* value there should be more total discrepancy from congruity toward the rating of the low *H* value word in the phrase than among phrases whose words differ less in *H* value. Item 3a in Table 3 indicates that this test of the hypothesis is supported by the data and is significant at less than the .025 level of significance. The second test of the hypothesis states that the total discrepancy from congruity toward the rating of the word with the higher *H* value should be less in phrases whose member words differ greatly in their *H* values than in phrases whose member words differ less in their *H* values. As Item 3b indicates, this test of the hypothesis is not supported by the data. The obtained difference is in the predicted direction but is not significant. The third test of the hypothesis states that the net discrepancy from predicted congruity ratings should be more toward the rating of the low *H* value word in the phrases when the words in a phrase differ greatly in their *H* values than when they are more similar. This test of the hypothesis is supported by the data and is significant at less than the .05 level of significance. Thus two of the three tests of the hypothesis support the general conclusion that discrepancies from congruity will be toward the rating of that member of a phrase which has the lower *H* value and that this tendency will increase as the difference in *H* values of the members of a phrase increases.

A brief comment must be offered concerning the failure of the second of the tests of the above hypothesis. Discrepancies from congruity toward the rating of the higher *H* value member of a phrase should have decreased as differences in *H* value between the words in the phrase increased. But this did not happen. To understand this failure, one must know two things about the Semantic Differential. First, since the congruity hypothesis weights the more polar word in a phrase more heavily, the predicted rating will be closer to the more polar word than to the rating of the less polar word. Second, the nature of the scales on the Semantic Differential restrict the possibility of giving a word a highly polar rating. It is impossible to rate a word higher than 3. Thus the distance between the predicted congruity rating and the upper end of a scale on the Semantic Differential will tend to be small. Any discrepancy from the predicted congruity rating toward a more polar rating must then be small. But such a discrepancy must take place if discrepancies toward the word with the higher *H* value are to occur. It has been shown in Hypothesis 1 that high *H* value words are more polar words. This means then that discrepancies toward the high *H* value word are restricted in their potential magnitude. They

will be small no matter what other conditions exist in the phrase. Due to the nature of the Semantic Differential, it would be virtually impossible to get a large discrepancy toward the high *H* value word in a phrase since it is the more polar word as well. Thus irrespective of the differences in *H* values of the words in the phrase, the total discrepancy from predicted congruity ratings toward the higher *H* value word would be small. It is most likely that this is the reason for the failure of the test of our hypothesis. It is due to an artifact of the measuring instrument in question.

There are, then, two major findings of this study. The first is that words and phrases with higher *H* values will be given more polar ratings on the Semantic Differential. A high *H* value implies that a word has many different free associations and that these associations are relatively equiprobable, while a low *H* value implies few different free associations, each with different probabilities of occurrence. Noble (1952) assumes that the number of free associations to a word is an index of the meaning of the word, with more associations indicating more meaning. Therefore, following his thinking, *H* should be an index of meaning since it describes the number of associations to a word with the additional virtue of taking into account their probabilities. If *H* indicates the amount of meaning of a word, then *H* should relate to the Semantic Differential ratings of the word. Polarity on the Semantic Differential is an Index of the amount of meaning of a word since it reflects the distance from a neutral or "no meaning" rating. Therefore, polarity should be the aspect of Semantic Differential ratings which correlates with *H*. This assumption has been validated in this study.

The second major finding showed that the *H* values of the words in a phrase were related to the Semantic Differential rating of that phrase. As the *H* values of the words in the phrase grew more different, the predicted congruity rating of the phrase was found to be more erroneous, and the discrepancy from the predicted ratings was in the direction of the word in the phrase with the lower *H* value. The question which needs to be answered is why the lower *H* value word should be of such importance in determining the rating of the phrase.

To answer this question it is necessary to consider what aspect of a word the *H* statistic may be measuring. It seems likely that this aspect may be ambiguity. A high *H* value word has many different associations which are more or less equiprobable. In a sense, it could be said that the high *H* value word has many different meanings which could occur fairly often. The low *H* value word, on the other hand, has few associations, and of these, only few occur with any great probability. The remaining associations are quite rare with respect to the rest. In other words, the low *H* value word has few different meanings, and of these only a small fraction occur with any considerable frequency.

Looking at the *H* statistic from this point of view, it seems reasonable to suppose that it may measure the ambiguity of a word. The common sense definition of ambiguity would state that an ambiguous word has many meanings *and* that these meanings occur often. If a word has several meanings but only one is common, it is not a very ambiguous word. For example, consider the word "mean." This word has at least three meanings of definitions: as a verb meaning "to refer to something," as an adjective meaning "cruel or bad," and as a noun meaning "arithmetic average." For the lay person the first two definitions may be considered ambiguous because they are different and common. The third definition produces no ambiguity for the statistically naive person because he is not aware of it. It has a low probability of occurrence. Therefore, we may consider the number of definitions of a word and their probability of occurrence as the determinants of the ambiguity of the word. It is assumed here that free associations are an indicator of the various meanings of definitions of a word. Since these "definitions" can be described by means of the *H* statistic, it follows that the *H* statistic gives us an indication of the ambiguity of the word. The greater the *H* value, the more the ambiguity.

If one accepts the assumption that the *H* value of a word is an indicator of its am-

biguity, it is not difficult to explain the above results. It would seem logical to assume that the less ambiguous word in a phrase is going to be an important factor in determining the reaction to the phrase. In this case the "reaction to the phrase" is a description of it on the Semantic Differential. The reaction to a two-word phrase can be considered a sort of choice reaction. One or the other of the words must be selected as the major determiner of the reaction to the phrase. Given a choice between two possible situations, it is likely that most people would choose the less ambiguous one. The less ambiguous one implies fewer possibilities for behavior, thus it could be understood and reacted to more easily. So in the case of phrases, it would be the less ambiguous word which would be "chosen" as the most important determiner of the reaction to the phrase. Its "meaning" is clearer and the reaction to it easier to make.

The congruity hypothesis as outlined by Osgood et al. (1957) does not take account of the ambiguity of the words used. It is therefore not surprising that their hypothesis should lack accuracy when words of different amounts of ambiguity (different *H* value) are used in a phrase. As it was shown in this study, errors in prediction from the congruity hypothesis tend to be in the direction of the less ambiguous word, that with the lower *H* value. It should be noted that these results do not in any way invalidate the congruity hypothesis. They merely suggest an additional bit of information which must be taken into account if one is to make accurate predictions about the Semantic Differential ratings of a phrase from the ratings of the individual words. In addition, the study suggested that in any use of the Semantic Differential, whether in the laboratory or clinical situation, a knowledge of the *H* values of the words or concepts rated is essential for better interpretation of the Semantic Differential profiles. It may be said that the implications of the study by Ross and Levy (1960) were substantiated. The *H* values of words have a bearing on their description on the Semantic Differential.

SUMMARY

The *H* values of words were determined by measuring the probabilities of free associations to the words and entering them into the equation for the *H* statistic. The words were rated on the Semantic Differential. Adjective-noun phrases were constructed by pairing each rated adjective with each rated noun. These phrases were rated on the Semantic Differential. In all, 108 subjects and 12 stimulus words were used.

1. The polarity of a word or phrase on the Semantic Differential was high when the *H* value of the word or phrase was high and low when the *H* value was low.

2. Using the congruity hypothesis, the ratings of the phrases were predicted from the ratings of the individual words on the Semantic Differential. These predictions became less accurate as the words in the phrase had more different *H* values.

3. The discrepancy from the predicted congruity rating of a phrase on the Semantic Differential tended to be in the direction of the word in the phrase with the lower *H* value.

These findings can be understood in terms of the hypothesis that the *H* statistic describes the ambiguity of a word. The high *H* value word, having more meanings and hence being more ambiguous, will also have more meaning on the Semantic Differential; a greater number of meanings of equal strength will appear as greater polarity or intensity of meaning on the Semantic Differential. The less ambiguous word having a more unitary meaning, will determine the rating of a phrase by virtue of the stability of its meaning, despite its lack of intense meaning or polarity.

REFERENCES

ATTNEAVE, F. *Applications of information theory to psychology*. New York: Holt-Dryden, 1959.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

ROSS, B. M., & LEVY, N. A comparison of adjectival antonyms by simple card pattern formation. *J. Psychol.*, 1960, **49**, 133–137.

(Received May 3, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 95–103

COGNITIVE BALANCE AND SELF-RELEVANT ATTITUDES[1]

MARC PILISUK[2]

Oberlin College

Among the many recent theories which apply to compatibility among cognitive elements, there is a high degree of agreement that certain attitudes or cognitions, when associated with one another are dissonant or in a state of imbalance (Abelson, 1959; Cartwright & Harrary, 1956; Festinger, 1957; Heider, 1944; Newcomb, 1953; Osgood & Tannenbaum, 1955; Taguiri, Blake, & Bruner, 1953). Perhaps the instance of imbalance most frequently sighted is that in which ego likes or respects both another person and an object or idea, while the other person dislikes that object or idea. Ego, in this instance, is considered by Heider's theory, for example, to be under tension which might be resolved by several alternative courses. The alternatives cited are: a shift in attitude toward the person, a shift in attitude toward the object, a cognitive separation of the person from the object, or action which in some way aims to alter the discrepancy; e.g., convincing the friend to change his mind (Heider, 1944, 1958). Moreover, in Heider's formulation of the problem, an additional alternative, involving no change but the perpetuation of a state of tension, was offered, and clinical documentation of emotional ambivalence seems to require that this alternative also be taken into account. An obvious difficulty with the formulation of this theory is that from a state of imbalance almost anything which follows may be "consonant" with the theory. However, certain modifications in the theory can be made to enable predictions from among the various alternatives.

The addition to balance theory being suggested is that there exists a dimension of resistance to change among attitudes. At the highest point of this dimension is the positive attitudes toward one's self, closely followed by attitudes toward a set of attributes, abilities, and ideas with which the self is identified; i.e., attitudes which support self-acceptance. Sharing second place in this dimension are attitudes toward a small group of significant others.[3] Imbalanced situations among these entities are termed central and all other attitudes derive a degree of centrality from the strength of their associations with members of the central group. Crucial to this conception is the hypothesis that attitudes in the central group, regardless of whether they are in balanced or imbalanced situations, rarely change their positive signs. When the imbalance obtains between two entities within the central group, the alternatives remaining are continuous tension or anxiety,[4] on the one hand, and a cognitive separation of the two entities, or rationalization, on the other. The experiment which follows attempted to test the validity of this conception.

METHOD

In that portion of the experiment to be described here, the subjects were 88 male college students from the University of Michigan.[5] The entire experimental group of 47 as well as 16 of the 41 control subjects were members of fraternities while the remaining 25 controls were drawn from a pool of students taking introductory psychology sections.

Subjects participated in groups of five. The method consisted of having a subject, in the role of group spokesman, perform a task which was presented to him as a measure of his personal competence. The group members' roles were ostensibly to provide the subject with critical evaluation of his performance.

[1] This research was supported under a predoctoral research fellowship, MF-9041, from the National Institute of Mental Health, United States Public Health Service. The writer is indebted to T. M. Newcomb and Daniel Katz for their assistance.

[2] Now at the University of Michigan.

[3] The significant other may better be defined as a member of ego's normative reference group (Kelley, 1952). It should be noted that neither the significant other nor the abilities and attributes referred to need correspond to real persons or to objectively measured qualities. The theory deals only with the attitude held toward such entities from the viewpoint of the individual person.

[4] Anxiety has been defined as a threat to the self-concept (May, 1950).

[5] Two additional subjects gave data which was discounted because of irregularities in the experimental procedure.

In actuality, the evaluation was arranged prior to the session by the experimenter and each participant was made to believe that he alone was the spokesman.

The task proper consisted of writing recommendations which might prove helpful in a casework report which was distributed. Each participant worked on this task independently. It was emphasized that the task required no special knowledge and was a measure of an individual's maturity, social sensitivity, and practical intelligence; attributes which can be regarded as supporting self-acceptance among a college population. It was further emphasized, for the purposes of strengthening ego envolvement, that the spokesman's final performance, after benefiting from the group's critique, would speak for the group as a whole in competition for prizes with other groups.

Subjects wrote their recommendations with four carbon copies. The avowed purpose of this duplication was to provide each critic with a copy of the spokesman's recommendations and to allow critics to see one another's recommendations. Recommendations were actually written in three stages, permitting a round of notes of criticism after each stage.

In this study an experimental subject is defined as one for whom one of the other four participants was a close mutual friend, i.e., each had rated the other as a high sociometric choice,[6] and the other three participants were strangers. A control subject is defined here as one for whom each of the other four members of the group were strangers. The sociometric indices were part of a battery of eight tests which had been administered to 154 members of six fraternities at a time several weeks prior to the experimental sessions and experimental subjects tended to regard the presence of a close friend as an accident.

Subjects were given an opportunity to memorize each others' names as part of the introductory procedure. The five roles, Critic Red, Critic Yellow, Critic Green, Critic Blue, and Spokesman were to be assigned after each person took his seat about a hexagonal Bavelas board. It was announced that critics would make their criticisms by sending a select group of typewritten statements from a large standard set made available. It was explained that the experiment, as a simulation of politician-constituent relations, required that the spokesman be able to differentiate his critics but not to have direct contact with them which would reveal their identities. In addition, it was asserted that the spokesman would be given more information needed for the task than would his critics so that the critics' evaluation would have to be based partly upon their judgment of the spokesman's competence and not merely on whether the spokesman's performance corresponded to their own.

One further procedural point should be mentioned here. Critics were instructed to send their notes of evaluation to the switchboard (the experimenter's booth of the Bavelas board) where the order of notes might be recorded before sending them on to their intended destinations. This detail is highly important for it made possible the use of false feedback. As told to the subjects, one could identify the source of a criticism note by the color of the paper on which it had been typed. In actuality, all criticism originated at the experimenter's booth. In fact, each subject was secretly assigned the role of Spokesman and the role of all four critics was played by the experimenter. The guise was maintained by a careful synchronization of instructions to critics and to the spokesman during the course of the experiment. Regardless of the subject's actual performance he received four standard sets of criticisms, each note identifiable as to source by color alone.

The subject received eight notes in each of the four colors. The notes were so constructed as to give the subject the impression that Critics Red, Yellow, and Green were accepting his performance in degrees varying from wholehearted acceptance to tentative or partial agreement, but that Critic Blue was being both harsh and abusive in his criticism. This impression of Critic Blue's distasteful evaluation was furthered when the illusory critics were asked to fill out a final evaluation of their spokesman. The experimenter then passed on to the spokesman information that Critic Red had rated his performance excellent, Critic Yellow, good, Critic Green, fair, and Critic Blue, extremely poor.[7]

The spokesman, or spokesmen to be more accurate, then filled out the first of two questionnaires. The questionnaires purported to measure the ability of the group to work well together. Page one of every questionnaire was a decoy addressed to the illusory critics. The part of the questionnaire actually used contained an introductory preface and a set of questions which were answered by placing checks along a nine-point rating scale labeled at the extremes.[8] Subjects checked each item four times

[6] In the sociometric index each subject selected his six "closest" fraternity brothers and ranked them according to the strength of his liking for them. The mean sociometric ranking of the friend among experimental participants was 2.1 with 1 being the person liked most, 2 next, etc. Subjects were also asked to rate these persons on a scale of how well each person chosen knew the subject. However, only the original rankings were used in the selection of mutual friends.

[7] The rating of extremely poor had been defined as follows: The spokesman is showing immature and rather foolish judgment. His recommendations are rather inadequate to speak for the group.

[8] The entire questionnaire is a modification of one used by Harvey (1958) in a similar study. For Items 2 through 10 of the questionnaire, directional predictions were made in keeping with the theory that persons would tend to rationalize the negative criticism. The predictions corresponded to differences between the answers given for Critic Blue and

selecting the appropriate color crayon to indicate the critic to whom each check referred. The following items were used:

1. On the average, how closely do you think the other person's recommendations agree with your own?
2. How serious do you think the other person was in the rating he made of you?
3. How socially sensitive is the other person who made the ratings of you?
4. How closely did the other person who rated you follow instructions?
5. Do you think the other person may have been angry or annoyed at you for some reason when he rated you?
6. Do you think the other person interpreted his job as critic and rater the same way as you would?
7. How careless was the other person in reading the case or in rating you?
8. Do you think that the other person would be able to give a legitimate defense of his opinions?
9. How easily do you think you could resolve any differences of opinion with this person if given a chance to talk together?
10. How closely do you think this person's appraisal of your character and ability would agree with your own appraisal?
11. Do you think that this person was rating you as a person or was he considering only what you wrote?
12. How adequate do you feel your original recommendations were for the case under consideration? (one answer)
13. How strongly do you desire to make any major change in your original recommendations? (one answer)
14. Guess the names of each of your critics.

To summarize the procedure to this point, each subject engaged in a task designed to reflect important attributes of his self-concept. He received four sets of anonymous evaluations, one of the four being highly negative and abusive. He then rated his critics on a variety of qualities. So far experimental and control subjects alike were involved in a situation in which one central attitude (toward the performance) and one peripheral one (toward an anonymous colleague's criticism of this performance) stood in an imbalanced relationship.

The last part of this experiment involved giving the subject information concerning the identities of his critics. The identity to be revealed for each color had actually been assigned prior to the experimental session. Assignments were made on a chance basis with one exception; i.e., for all experimental subjects, the highly rated mutual friend was revealed as the negative critic, Blue. Actual revelation was gradual, with subjects first given only ordinal information placing the person assigned to be Blue as a critic who was less favorable than one of the other group members. The name of each critic was then given to the subject and a final questionnaire was distributed. A recall task was added to the former questionnaire as were the following three items:

14a. How well does the other person who rated you know you? Using 10 to stand for the person who knows you best (whoever it is) rate the other person from 0 (doesn't know you at all) to 10, depending on how well you think he knows you.

14b. How sure are you that the other person actually wrote the ratings of you that you saw?

14c. How sure are you that the other person actually saw your recommendations?

The remaining items were identical to those answered on the first questionnaire.

In summary, this last stage placed the experimental subject in a situation involving imbalance among central entities. A "significant other" had been revealed as the critic of an attribute one's self-concept.

RESULTS

The primary results deal with resistance to change among central attitudes. The first of these central attitudes is the positive attitude held toward one's performance in a task said to be a measure of one's self-competence. The results are shown in Table 1.

None of the differences in Table 1 is statistically significant for either item (using a t test for related samples). In both experimental and control conditions the mean rat-

TABLE 1

SATISFACTION WITH CRITICIZED PERFORMANCE OF SELF-RELEVANT BEHAVIOR

Item	Before knowledge of critic's identity		After knowledge of critic's identity	
	Experimental group[a]	Control group[b]	Experimental group[a]	Control group[b]
How adequate were your recommendations?[c]	6.6	6.3	6.8	6.5
How strongly do you desire to make changes?[d]	2.0	2.3	2.1	2.4

Note.—Mean ratings on a nine-point Likert-type scale from 0 to 9.
[a] $N = 47$.
[b] $N = 41$.
[c] High scores indicate very adequate.
[d] High scores indicate strong desire to change.

those given for the median positive critic prior to identification of the critics. The predicted directions may be seen in Table 4. Directional predictions regarding changes in the rating of Critic Blue by experimental subjects after identities of the subjects were known, appear in Table 6. The explicit statement of all hypotheses tested in this study may be found in Pilisuk (1961).

TABLE 2

PERCEIVED STRENGTH OF ACQUAINTANCE BETWEEN CRITIC AND SELF

Group	Mean rating of positive critic[a]	Mean rating of negative critic
Experimental ($N = 47$)		
Pre-experiment sociometric	not obtained	7.3
Postidentification rating	1.6	8.3
Control ($N = 41$)	5.1	1.3

[a] Ratings go from 0 (doesn't know me) to 10 (knows me best).

ing for adequacy of own performance was significantly higher than the midpoint of the scale ($p < .001$ using the sign test, one-tailed). Also in both conditions the mean rating of desire to make changes in the performance was significantly lower than the midpoint of the scale ($p < .001$ using the sign test, one-tailed). Thus subjects tended to maintain satisfaction in their performances despite the presence of an abusive critic of this performance. This tendency was equally strong among those subjects for whom the negative critic was identified as a close friend.

The next attitude which has been called central is a positive attitude toward a friend. The item used to measure this attitude is, "How well does the other person who rated you know you?" Data obtained on this item are listed in Table 2.

The most relevant comparison is between the responses of the experimental group to the friend on a sociometric inventory prior to the experiment and the response to this same friend after receiving an insult from him. The strength of acquaintanceship was deemed slightly higher after the friend's insult than before it, the difference being significant at .05 level, two-tailed. Certainly the estimation of acquaintanceship did not diminish.

One might be inclined to infer at this point that the experimental subjects were looking upon the entire situation as a hoax. But data on the credibility items (see Table 3) do not support this view.[9]

None of the differences between control and experimental groups in Table 3 is statistically significant, and none of the mean ratings of credibility falls below the midpoint on the scale. In this experiment subjects tended toward a belief that the circumstances presented were true.

The results above should not be taken to suggest that subjects stood by and good-naturedly accepted abuse. While neither friendship nor self-competence was permitted to decline, reactions to the negative critic served to rationalize the harsh criticism and thereby destroy its validity. Table 4 refers to the reactions by all subjects to their, as yet, anonymous critics. Predictions were made for all items in which rationalization could be inferred by the direction of the differences.

In the case of negative information received from an anonymous source, subjects tended to make use of all available rationalizations. Some of these may be looked upon as an attempt to cast the negative critic in an undesirable mold. This is one means of restoring consonance to the situation at hand. No direct measure of dislike for the harsh critic was attempted. Correlations among the various items of rationalization ranged from .39 to .73 indicating a tendency for those subjects who found one source of blame for Blue's hostile criticism to find other sources as well. That is, they formed a negative stereotype of this critic. Those who revealed less discrepancy between the characteristics attributed to the negative critic and those

TABLE 3

CREDIBILITY OF POSITIVE AND NEGATIVE CRITICISM[a]

Item	Experimental group ($N = 47$)		Control group ($N = 41$)	
	Negative critic	Median positive critic	Negative critic	Median positive critic
How certain are you that other person wrote the criticism?	4.7	6.1	5.0	6.8
How certain are you that other person saw your work?	5.1	6.3	4.7	6.4

Note.—Mean ratings are on a nine-point scale from 0 (very doubtful) to 9 (quite certain).

[a] All ratings were obtained after knowledge of critics' identities had been revealed.

[9] Disbelief was actually a common occurrence in several pilot attempts. Great effort was put into the achievement of credibility in the final experimental situation.

TABLE 4

MEAN RATINGS TO ANONYMOUS SOURCES OF CRITICISM

($N = 88$)

Topic of item	High scores indicate	Mean rating given blue (B)	Mean rating given median (M)	Direction predicted
How closely does person agree with you?	Completely	2.4	7.9	None
How serious was the other person?	All in fun	3.0	1.7	B > M[a]
How socially sensitive is the other person?	Extremely insensitive and socially blind	6.2	2.2	B > M[a]
How closely did the other person follow instructions?	Followed them completely	5.8	6.9	M > B[a]
Do you think the other person may have been angry or annoyed at you?	Positive he was angry	2.9	1.0	B > M[a]
How did the other person interpret his role as critic?	Exactly the same as I would	4.2	6.4	M > B[a]
How careless was other person?	Very careless	4.7	2.2	B > M[a]
Could other person give a legitimate defense of his criticism?	Definitely	3.6	6.9	M > B[a]
How easily could you resolve difference of opinion if you could talk together?	Very easily	3.0	7.2	M > B[a]
How closely would this person's appraisal of your character agree with your own?	Very close agreement	1.8	6.4	M > B[a]
Was person evaluating only your behavior or only you as a person?	Only as a person	2.8	2.2	None

Note.—Critic Blue is a source of harsh negative criticism. Critic median refers to the median rating between three sources of positive criticism. Ratings were made on a scale ranging from 0 to 9.

[a] Difference significant at .001 probability level (one-tailed).

attributed to positive critics tended to minimize the discrepancy consistently.

After the names of the critics had been revealed, the control subjects still remained in much the same situation as before. The anonymous critics now bore names but the critics were still strangers. It is interesting that in this condition the control subject tended to attribute knowledge of himself to the more favorable critics while denying it to the negative critic (see Table 2).

There is some indication in the data that the identification of the close friend as the source of negative criticism was resisted by the experimental group. After receiving a clue that the person assigned the role of Critic Blue had been a less favorable critic than one other person, maximal use of this clue would demand a guess that the less favorable of these two critics was indeed Critic Blue.

The clue was standard for experimental and control subjects. The roles of the critics were assigned Numbers 1 through 4 (red, yellow, green, and blue in that order) in accordance with the degree to which criticism received from them had been of a positive nature. Maximum use of the clue calls for assignment of the less favorable critic to a rank of 4; i.e., a guess that he is Critic Blue. Minimum use would call for assigning the rank of 2; i.e., a guess that he is Critic Yellow. The control group for whom no friend was involved obtained a mean score of 3.7. The experimental group for whom Critic Blue was a close friend obtained a mean score of 2.7. This difference is significant at $p > .001$. Experimental subjects appeared to make less efficient use of the clue (see Discussion).

The identification of the critics brought about a major shift among the experimental group from the situation existing prior to identification. For these persons a close friend was identified as having criticized a salient aspect of the individual's self-concept. As already shown, neither the assessment of criticized performance nor the assessment of the acquaintanceship tie decreased as a consequence of the new situation in which the friend had turned into a harsh critic. The responses which did occur were in keeping with expectations from the theory of balance.

Upon receiving notes revealing identities of

TABLE 5

CHANGES IN JUDGMENTS ATTRIBUTABLE TO THE NEGATIVE CRITIC'S IDENTIFICATION AS A CLOSE FRIEND OF THE RECIPIENT

Focus of change	Mean amount of change[a]	
	Experimental (N = 41)	Control (N = 41)
Change toward greater perceived agreement between the recipient and the negative critic	.64	.00
Total change in attributions to a negative critic[b]	16.9	8.0
Total change in attributions to three positive critics[b]	23.1	15.0

Note.—Significance of difference between groups: $p < .001$ in all cases.
[a] Change measured in scale units.
[b] Summed over Items 2 through 11 on experimental questionnaires.

the critics, a number of experimental subjects revealed stress in a variety of ways. Some shook their heads, others glared at the notes which contained this information, some laughed aloud or sighed, and a number were quite obviously blushing. There seemed also to be a greater time lag before they got back to work. But none of these indications of tension was recorded systematically. What was recorded were the changes in a subject's questionnaire ratings from the first to the second administration.

Unlike the preidentification condition, experimental subjects in the postidentification situation were highly selective in the items chosen to account for the friend's misdeed. While one subject might change his view to suggest that he now thought the negative critic was angry, another would modify his view to favor jesting as the rationale. Despite the fact that individual differences were wide, certain regularities were observed. These may be evaluated by comparison with the control group and are presented in Table 5 below.

The results may also be evaluated by using each individual as his own control; i.e., by comparing preidentification and postidentification questionnaires and testing for significant changes made in the reactions to the negative critic. These are shown in Table 6.

The results in Tables 5 and 6 depict the consequences of bringing into an imbalanced configuration two cognitive entities neither of which would change. The first item in Table 5 demonstrates a reliable tendency to reduce the size of the perceived incompatibility. Although neither friend nor subject had opportunity to change their performances, the

TABLE 6

THE EFFECTS OF LEARNING THAT A NEGATIVE CRITIC WAS ONE'S CLOSE FRIEND UPON CHARACTERISTICS ATTRIBUTED TO THIS CRITIC

(N = 47)

Topic of item	Direction of difference from first rating	Mean change in scale units
How closely does person agree with you?	More closely[a]	.64****
How serious was the other person?	Less serious[a]	1.32****
How socially sensitive is the other person?	More sensitive[a]	.81****
How closely did the other person follow instructions?	Less closely[a]	1.00**
Do you think the other person may have been angry or annoyed at you?	More angry[a]	.21
How careless was other person?	Less careless[b]	.45
Could other person give a legitimate defense of his criticism?	Less legitimate[a]	.43*
How easily could you resolve difference of opinion if you could talk together?	More easily[a]	.94***
How closely would this person's appraisal of your character agree with your own?	More closely[a]	1.55****
Was person evaluating only your behavior or only you as a person?	More as a person[a]	.17

[a] Predicted direction.
[b] Opposite direction from that predicted.
* $p < .05$.
** $p < .01$.
*** $p < .005$.
**** $p < .001$.

subject now sees the area of disagreement to have shrunken slightly. On the remaining items of the questionnaire experimental subjects made significantly more shifts (rationalizations) than did the controls. It is interesting to observe here that the experimental subjects also made significantly more changes in their reactions to the strangers, who had rated them favorably, than did the controls. The indication is that when one finds himself rationalizing in order to exonerate the friend the actual scale on which the evaluations are being made undergoes a change in meaning.

As mentioned before, group differences on individual items tend to mask the high degree of individual variability among experimental subjects' responses to the postidentification situation. The standard deviations of the experimental group's scores, on every item, increased from first to the second questionnaire and full accounting for the individual differences found requires more knowledge of personality characteristics of individual participants. Such considerations lie beyond the scope of this paper.[10] However, several changes significant for the experimental group as a whole are seen in Table 6. The largest changes were found on items suggesting that the friend had been less serious than originally assumed and that the friend's opinion of ego's personality was in closer agreement with ego's opinion than had previously been assumed.

DISCUSSION

The data make possible an addition to current theories of attitude change. Simply stated, the addition is that attitudes toward persons or entities, which are prized, tend not to change even when such change might serve to reduce dissonance or tension. This finding is noteworthy because of its apparent consistency with other explanations of the importance of anchorage points in judgments about new information (Helson, 1947; Sherif & Sherif, 1956). Just as the extreme weights in an assortment of objects determine our judgments of the weight of intermediate objects, so too, it would seem, do a limited number of highly valent attitudes tend to remain highly valent and establish the frame of reference by which new information is evaluated. In this study subjects were first confronted with two pieces of incompatible information. The first piece of information was that they had performed a task which purported to test their personal competence. Second, they were told that an anonymous colleague disdained this performance. In consequence, the negative critic was himself rejected, while the more central anchorage point, the competence testing performance remained positive. In the case in which two stable anchorage points, self-competence and a close friend, were placed in an imbalanced relationship, a further change; i.e., rationalization, was used to permit the maintenance of a high positive valence for both points of anchorage.

The evaluative dimension is the dimension most frequently found in factor analytic studies of connotative meanings of words (Osgood & Tannenbaum, 1955). This magnifies the importance of a definition of such evaluative anchorage points as the attributes of one's self-concept and the allegiance of significant others. In the reception of social information, such anchorage points may have as pervasive an effect upon our judgments as the horizon and the vertical plane may have upon our perception of space.

The finding reported in Table 2 which showed a tendency among control subjects to attribute knowledge of self to the strangers who praised their performances is consistent with the interpretation. But note also that the experimental group did not grant this degree of acquaintanceship to the supportive group members. For the experimental subject to do so he would have had to interpret the questions in a way which would have reduced his estimation of another stable anchorage point, i.e., his close friend. It would seem that attribution of acquaintanceship to flatterers is desirable, but not at the expense of the alienation of friends.

One interesting finding was that estimations of acquaintanceship by friends appeared to increase after experimental subjects had received an insult from these friends. Dissonance theory would lead one to expect that a negative experience from a close friend

[10] Data relevant to considerations of individual differences were obtained and are being prepared for publication.

could be balanced by an intensification of one's sense of friendship. However, some caution should be taken in interpreting the data in this way for the difference found may be an artifact of the measures used. When a subject first rated the friend as to degree of acquaintanceship it was in a context in which he was rating six close friends, i.e., as part of the sociometric index. The second rating occurred in a context in which the friend and three strangers were being rated. It is quite conceivable that, contrary to instructions, some subjects were making relative ratings which diminished the friend rating in the pre-experiment situation and favored the friend in the postidentification period.

The finding that experimental subjects did not guess the friend at as low a point as maximum utilization of the clue given to them demanded should also be interpreted cautiously. Even some of the control subjects made less than maximal use of the clue to the negative critic's identity. In addition, if one takes into account the initial guess as to identities of the critics (before a clue was given), then the experimental group moved, in the direction suggested by the clue, an amount almost equivalent to the control group. Further ambiguity in the finding is brought about by the fact that the guesses of the control group approached a fixed limit. The control group's utilization of the clue might have been further distinguished from the experimental group if the limitation were removed.

Support for the hypothesis that attitudes defined as central tend to resist change suggests the interpretation that the stability of significant others and of self-attributes provide landmarks in a person's evaluative map of his environment. To postulate that a need for such stable landmarks exists should not imply that the need is always fulfilled. Neither does the hypothesis of a tendency to resist change mean that persons invariably retain their friends or always look with favor upon their own abilities. In this experiment, the subject was reinforced in assessing the competence of his performance by positive evaluation from several critics and by assurance that he had sounder preparation for the task than had any of his critics. In addition, several alternative explanations were available to permit the subject to exonerate the friend for criticism attributed to him. The point to be emphasized is that where alternatives to change in central attitudes are possible, they will be selected.[11]

This experiment highlights a need in future research, to measure those attitudes which are most central to an individual's conception of himself. McClelland (1955) has argued the case for a study of mental content. What content could be more in need of precise assessment than those attitudes which people hold too dear to permit changes?

SUMMARY

An experiment was performed to test the hypothesis that attitudes toward attributes of one's self-concept and toward significant others tend to resist change. The hypothesis was presented as an addition to the theory of cognitive balance. The method consisted of a group experiment in which a subject performed a task which was alleged to be a measure of his personal competence. The ostensible function of the group members was to give the subject a critical evaluation of his performance, in particular, and his competence in general. A system of restricted communication and false feedback permitted the experimenter to play the roles of all critics, one of whom sent negative and abusive criticism. In the experimental group a close friend (high sociometric choice) of the subject was alleged to be the source of the adverse criticism. In the face of adverse criticism, whether from friend or stranger, the subject retained a favorable estimation of his performance. In the experimental group, estimates of the friend's acquaintanceship were not diminished by the friend's behavior. A wide range of rationalizations were used, however, to explain the situation. The interpretation of the findings was drawn to suggest that there exists a group of attitudes which provide stable anchorage points in the evaluation of new information.

[11] It is conceivable that a general level of dissonance among central entities is itself a personality characteristic which may determine one's ability to tolerate further changes among highly valued entities (Pilisuk, 1961).

REFERENCES

ABELSON, R. P. Modes of resolution of belief dilemmas. *J. conflict Resolut.*, 1959, **3**, 343–352.

CARTWRIGHT, D., & HARRARY, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, 1956, **63**, 277–293.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

HARVEY, O. J. Reactions to negative information about the self as a function of the unfavorableness of the information, source of the evaluations and personality characteristics of the recipient. Paper read at the American Psychological Association, 1958.

HEIDER, F. Social perception and phenomenal causality. *Psychol. Rev.*, 1944, **51**, 358–374.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

HELSON, H. Adaptation level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

KELLEY, H. Two functions of reference groups. In G. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (2nd ed.) New York: Holt, 1952. Pp. 410–414.

MCCLELLAND, D. C. The psychology of mental content reconsidered. *Psychol. Rev.*, 1955, **62**, 297–302.

MAY, R. *The meaning of anxiety.* New York: Ronald, 1950.

NEWCOMB, T. M. An approach to the study of communication acts. *Psychol. Rev.*, 1953, **60**, 393–401.

OSGOOD, C. F., & TANNENBAUM, P. The principle of congruity and the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.

PILISUK, M. Cognitive balance and the reception of unpleasant information: An exploratory study. Unpublished doctoral dissertation, University of Michigan, 1961.

SHERIF, M., & SHERIF, CAROLYN W. *An outline of social psychology.* New York: Harper, 1956.

TAGUIRI, R., BLAKE, R. R., & BRUNER, J. S. Some determinants of the perception of positive and negative feelings in others. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received May 24, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 104–111

THE EFFECTS OF REINFORCEMENT AND AROUSAL ON THE PSYCHOMOTOR PERFORMANCE OF CHRONIC SCHIZOPHRENICS[1]

ATHAN KARRAS[2]

University of Rochester

Two views have been advanced to explain the improvement of chronic schizophrenics on a simple psychomotor task when responding under an escape condition from an aversive stimulus. One group (Cohen, 1956; Rosenbaum, Grisell, & Mackavey, 1957) suggests that an aversive stimulus is a negative reinforcer, i.e., subjects respond as quickly as possible to escape it. The alternate explanation, introduced by Lang (1958, 1959), is based on stimulus dynamism and its effect on the arousal level of the organism. The improvement manifested by chronic schizophrenics exposed to intense excitation "may reflect a difference in the response of normals and schizophrenics along the intensity dimension of the stimulus" (1959, p. 263) so that the dynamogenic properties of the high intensity stimulus "possessed enough intensity to 'pull' the response from the organism despite lack of readiness to respond" (1958, p. 58). It is assumed, states Lang, that the chronic schizophrenic may be at a lower level of arousal than normals who function nearly optimally under typical test conditions.

The experimental design (Lang, 1958, 1959) used to test Lang's hypothesis might have confounded reinforcing with arousal properties of a collateral intense stimulus. Three of his experimental conditions introduced high intensity noise during a visual two-choice reaction time problem. One of these conditions included an Escape condition, the use of which in previous experiments with schizophrenics made Lang question its reinforcing effect. In this treatment the subject escapes from the aversive stimulus which is presented with every reaction time trial by making the correct response which terminates the stimulus. A second condition, the Avoidance condition, more clearly tested the reinforcing effect of the aversive stimulus, for the subject can avoid completely the stimulus by responding at a certain minimum speed. Under this condition it is possible for the subject to receive only one presentation of the aversive stimulus to improve for the duration of the experiment. Clearly, if improvement is obtained without the presentation of the noxious stimulus on every trial, improvement may be attributed to the reinforcing effects of avoiding the stimulus. Another, his Excitation condition, was to test solely the dynamogenic effect by presenting a high intensity stimulus independently of the subject's response. If improvement resulted in both the Escape and Avoidance conditions, Lang (1959) would assume that the Escape condition had reinforcing properties; but if the Excitation condition along with the Escape condition showed improvement then it was to be assumed that the Escape condition had dynamogenic properties. The outcome was that the Excitation and Escape groups differed significantly from the Control group, i.e., they improved, whereas the Avoidance group did not differ from the Control group. The conclusion was drawn that for chronic schizophrenics an aversive stimulus had dynamogenic properties which aid the patients and their task.

However, for the Excitation group the collateral aversive stimulus commenced with the visual signal to respond and was terminated independently of the subject's response after either a .40 second, .80 second, or 1.20 second duration of presentation. If we consider

[1] This paper is adapted from a thesis submitted in partial fulfillment of the requirements for the PhD degree at the University of Rochester.

The author wishes to express his appreciation to Burton G. Andreas, John H. Flavell, and Benjamin F. McNeal, members of the doctoral committee, for ther valuable advice and encouragement during the course of the research.

[2] Now at Psychiatric Institute and Hospital, Columbus, Ohio.

the subjects were chronic schizophrenics and that the time of cessation of the collateral stimulus was very close to their RTs, we may consider it as very likely that the termination of the intense white noise was a reinforcing event for them (even though they had been told earlier that their responses did not control the aversive stimulus). Further, analysis of the data reveals that the Avoidance group was not significantly different from the Control group, the slowest, or from the Escape and Excitation groups, the fastest, which were different from the Control group, i.e., the Avoidance group fell in between these extremes, suggesting that some improvement did occur. The reason for its being suggestive may be attributed to the fact that it was the subject's own performance which set the level of avoidance, so that only a slight improvement in RT was necessary to avoid the aversive stimulus. Apparently, this slight improvement was not of sufficient magnitude to reach statistical significance.

Apart from the lack of an experimental paradigm which conclusively tests the hypothesis, the hypothesis as formulated by Lang (1958, 1959) raises some questions, only two of which will be reviewed. It is not certain that chronic schizophrenics are, as Lang says, at a lower level of arousal than normals. A perusal of Malmo's work (Malmo, 1957; Malmo, Shagass, Belanger, & Smith, 1951; Malmo, Shagass, & Smith, 1951) indicates that under normal resting conditions chronic schizophrenics are for many measures of arousal at the same level as normals; in addition, for some stresses these patients show greater physiological reactivity than normals, but never exhibit less physiological responsiveness, even if their overt behavior at times suggests it. These findings are contrary to those cited by Lang (1958, 1959).

Second, concerning dynamogenesis, it is difficult to see why the introduction of a collateral intense stimulus which has no reinforcing properties should "pull" the correct response from the organism to improve performance. If we adopt Hull's (1951) view of stimulus intensity dynamism we should then concern ourselves with the discriminating stimulus instead of the collateral one because for Hull dynamogenesis deals essentially with conditioned stimuli which are moderate in intensity. If we adopt an earlier view of stimulus dynamism, as Hebb (1955) does, i.e., the activating or general drive effect on the organism, then it is true that responses will occur more strongly when the intensity of stimuli is increased. If general drive or arousal is elevated however without reinforcing a preferred response, it is difficult to predict that this response will be evoked with greater strength, since the occurrence of other responses is also strengthened which may interfere with the preferred one.

The question is more clearly stated as whether the arousal property of an intense stimulus characterized as aversive is of itself sufficient to improve performance or whether such a stimulus must be employed as a negative reinforcer. It is difficult to infer from current theory and data that an increase solely in the subject's level of arousal improves the performance of chronic schizophrenics. It is possible for reasons other than we now know of that this might be the case or it is possible that a sheer increase in the level of arousal without any cue for appropriate action may deteriorate performance by introducing competing responses. The present study was undertaken to determine if chronic schizophrenics improve on a simple psychomotor task because of the reinforcing or arousal effects of an aversive stimulus. In addition, the effects of adaptation were studied. Last, the subsequent effects of these two factors were investigated in order to evaluate the extent to which modified behavior persisted.

METHOD

Subjects

The study included 80 white male chronic schizophrenics from the Veterans Administration Hospital at Canandaigua, New York, divided into five groups in the manner described below. An F test for homogeneity of variance and an analysis of variance for age, education, and length of hospitalization resulted in no significant differences among groups at $p < .25$. The combined groups mean for age was 36.70 years (range: 25–45 years); for education, 11.48 years (range: 8–20 years); and for hospitalization, 9.11 years (range: 3–18 years). In addition to a minimum of 3 years of hospitalization patients were diagnosed as chronic schizophrenics and considered as such by the ward physician. None

of the subjects displayed signs or had histories of neurological involvement or primary mental deficiency. None were lobotomized at any time or had received any kind of shock treatment or barbiturates for over a year. All subjects were on tranquilizers.[3]

Apparatus

The apparatus was designed for two purposes. One, the measurement of the subject's RT in response to visual signals in a two-choice RT task; the other, the collateral use of white noise for four of the experimental conditions.

For the measurement of RT, the subject was seated before a table on which a baseboard was placed. A rest button was affixed to the board directly in front of the subject and 10.5 inches from the press buttons of two telegraph keys which were mounted on the board. Fifteen and one-half inches in front of the telegraph keys, and attached to the baseboard, was a vertical panel 25 inches × 30 inches in which were placed three .5 inch jeweled pilot lights. The lights, on the same level with each other, were about eye-level for the subject. One of the lights, the ready signal, was centered and appeared green when illumined. Eight and one-half inches on either side of the ready light was a red response light. All three were adjusted to the same level of subjective brightness. A light pressure on the right telegraph key turned off the right response light and the left telegraph key the left response light. A Standard Electric timer, activated simultaneously with either response light and terminated by the subject's correct R, recorded RT to the nearest .01 second.

The sequence of the presentation of the duration of the interval between the ready signal and a response signal and the sequence of right and left response lights were determined from a table of random numbers, with the restriction that no interval or response light be presented more than three times in succession. The intervals were either 2, 3, 4, or 5 seconds. All subjects received the same sequences. The average duration of the experimental condition was 15 minutes, with an average of 5–6 RT trials per minute.

The white noise was recorded on tape at a constant level of 116 db from a white noise generator. Loops were made of the tape to fit around the spindles of a Wilcox-Gay Recordio-Imperial. With a loop only a few seconds of recorded sound provided a constant sound signal. The sound was delivered to the subject through a pair of matched earphones which were cushioned. A voltmeter was used to monitor the energy input into the earphones.

Procedure

The subject made a jump reaction from the rest button to the telegraph key under which the response light was illumined. Performance was measured in terms of the latency of the subject's R to the response light. Prior to participating in the task the subject was first familiarized with the apparatus and then introduced to the sound stimulus which was first heard through the speaker of the tape recorder, but which thereafter was heard through the earphones. In order to be included in the study, the subject had to tolerate a minimum high intensity sound. Four subjects did not meet this criterion. Such a sound level, which also had to be experienced as unpleasant, was first tentatively established by four staff psychologists at the hospital and pilot tested on several patients not used in the study. After it was determined that the subject could tolerate a very intense sound he was asked to judge a noise level "in between a loud and low noise, somewhere in the middle." In this way a medium intensity level was determined for 65 of the 80 subjects. The other 15 subjects, the last 15 of the study, and comprising the Low Escape group, selected a moderately low intensity level which, though distinctly audible, had no reported unpleasant quality to it when suddenly presented. (It was found that a medium intensity noise when suddenly presented had a sharpness to it, and that this new level with its sudden onset sounded as loud as the continuous medium intensity noise.) The voltmeter readings of the energy levels were later calibrated for acoustic output. The low intensity range was from 47–51 db; the medium intensity range from 79–82 db; and the high intensity range from 91–94 db.

Following these preliminaries, the subject was instructed in the use of the apparatus and was given a few practice trials. When it appeared he understood his task he was given a 1 minute break. After this break the pre-experimental condition began in which the subject was given 36 RT trials, the first 15 of which were not used in the analysis. During the course of the first 15 trials the subject was encouraged seven times to respond as quickly as possible. He was given a report on his performance for each trial and told to improve if he slowed down or told to try to outdo himself if he improved. The remaining 21 trials were administered without encouragement. The subject then received a 1 minute break and was assigned to one of five groups matched for RT on the basis of the best median for each subject in the pre-experimental

[3] One reason for using only tranquilized subjects was that they constituted over 75% of the patient population at the hospital, a finding typical of other NP hospitals. Also, studies by Monroe, Heath, Mickle, and Miller (1955) and Bruck (1958) demonstrate that EEG patterns of tranquilized subjects are similar to those not tranquilized. EEG patterns are a measure of the level of arousal of the CNS (Malmo, 1959). However, the evidence in this area is ambiguous. More pertinent are the findings of a host of studies, e.g., Whitehead and Thune (1958), which clearly indicate that the chronic administration of tranquilizers does not affect the psychomotor and cognitive performance of chronic schizophrenics. Thus in ways crucial for this study tranquilized subjects are similar to nontranquilized ones.

condition in which the last 21 trials were divided into three blocks of seven trials each with their medians computed during the break. Thus subjects were matched on their pre-experimental scores for all but the Control group, which at first had more subjects than the other groups. When the decision was made to reduce the N of the Control group to that of the others (see Results) a subjects' match was made on pre-experimental scores to correspond to the other groups.

After group assignment, the subject received 84 trials of his experimental condition. The five conditions, with their instructions, are as follows:

1. Medium Stimulation (MS), $N = 15$. The sound was presented at the subject's medium intensity level and was left on throughout the entire condition. The subject was instructed not to pay attention to the noise and to concentrate on reacting as quickly as possible.

2. High Stimulation (HS), $N = 15$. The sound was presented at the subject's high intensity level and was left on throughout the entire condition. The instructions were the same as for MS.

3. Low Escape (LE), $N = 15$. The low intensity sound commenced with the illumination of a response light and was terminated by the subject's correct R. The subject was told that the sound would stop when he responded correctly. He was told that the reason for having the sound was to remind him to react as quickly as possible.

4. High Escape (HE), $N = 15$. The high intensity sound commenced with the illumination of a response light and was terminated by the subject's correct R. The subject was told that the sound went off when he responded correctly. In addition, he was told that the sound might seem very unpleasant and that he controlled how long he had to bear the discomfort by how quickly he reacted.

5. Control (C), $N = 20$. The Control group was divided into two equal halves. One half wore earphones without receiving sound. Subjects were told that the earphones were to dampen outside noises to help them concentrate on reacting quickly. Subjects in the other half were told to continue as in the pre-experimental condition, always reacting as quickly as possible, even if the task seemed tedious to them.

At the end of the experimental condition the subject took off his earphones and received a 1 minute break. During the break he was instructed to continue for a few more trials just as he had for the pre-experimental condition and to concentrate on reacting as quickly as possible. After the break 28 post-experimental trials were administered, as in the pre-experimental condition.

It was determined that the correction method was to be employed in case the subject hit the wrong key. Only four instances of the wrong response were made.

RESULTS

For each subject RT trials were grouped into blocks of seven and for each block a median was obtained. Thus the subject had 3 median scores for the pre-experimental condition, 12 for the experimental condition, and 4 for the post-experimental condition. The overall group performance measure of a condition used in the data analysis was the mean of all medians of a group, referred to as the Overall Mean.[4] In order to test for the effects of an experimental condition over the course of the 12 blocks of a treatment, the mean of medians of a group was obtained for each of the 12 blocks of its experimental condition.

TABLE 1

OVERALL MEANS AND *SDs* FOR PRE-EXPERIMENTAL, EXPERIMENTAL, AND POST-EXPERIMENTAL CONDITIONS[a]

Group	Condition					
	Pre-experimental		Experimental		Post-experimental	
	M	*SD*	*M*	*SD*	*M*	*SD*
MS	64.91	11.90	74.51	13.93	71.02	14.21
HS	63.82	10.96	70.11	13.68	66.32	12.93
LE	64.47	13.03	63.26	13.98	67.15	14.12
HE	64.71	16.90	57.52	15.43	66.43	17.75
C	64.50	14.43	64.99	12.34	64.18	11.04
Combined groups	64.39	13.60				

[a] Data are in units of .01 second.

Pre-experimental Performance

The groups did not differ in mean RT or variability for the pre-experimental performance (Table 1). All groups were positively skewed. It was found that both halves of the Control group were matched with Overall Means of 65.82 and 64.37 hundredths seconds and *SDs* of 12.01 and 15.23 hundredths seconds (for equated groups: $t = .223$, $p < .80$, *ns*).

[4] In preliminary analysis of the data the Best Mean of a condition was calculated, based on the mean of the best medians of subjects in a group. This mean was used for comparability with Lang's (1959) data which were derived from the subject's best performance. Since the Best and Overall Means in the present study resulted in almost similar findings, the customary Overall Means are presented in the text. The one exception to similar findings was the LE group which, with the Best Mean, had a slightly faster RT mean during the experimental condition than the Control group.

Experimental Performance

Before comparisons were made between the five groups on their experimental scores, a t test for equated groups was performed to test the significance of difference between the two halves of the Control group; the obtained t of .363 was not significant at $p < .70$. The halves were combined and five subjects were discarded to bring the final N to 15, while maintaining a subjects' match on pre-experimental scores. The pre-experimental, experimental, and post-experimental means and SDs reported for the Control group are based on these 15 subjects.

It was then found for the experimental condition that although homogeneity of variance was evident from inspection (see Table 1), the forms of the distribution were heterogeneous. Under such conditions the transformation of raw scores is not a useful technique (Lindquist, 1953). Nonparametric statistics were utilized to test for differences among the Overall Means of the experimental condition. A Friedman two-way analysis of variance was performed and was found significant at $p < .01$ with a $\chi^2_r = 45.79$. Wilcoxon matched-pairs signed-ranks tests were then performed to test the difference between individual pairs of Overall Means. The results are presented in Table 2. In brief, all but the LE group were significantly different from the Control group. The Overall Mean of the HE group was the fastest, while the two stimulation groups had the slowest Overall Means, with the MS group having a larger RT mean than the HS group.

TABLE 2

WILCOXON MATCHED-PAIRS SIGNED-RANKS TESTS BETWEEN GROUPS IN THE EXPERIMENTAL AND POST-EXPERIMENTAL CONDITIONS

Condition	Group			
	HE	LE	HS	MS
Experimental				
C	15.0[a]	31.0	1.5[a]	8.0[a]
MS	1.0[a]	2.5[a]	6.0[a]	
HS	1.0[a]	7.0[a]		
LE	2.0[a]			
Post-experimental				
C	18.0	0.0[a]	15.0[a]	6.0[a]
MS	20.0	3.0[a]	2.0[a]	
HS	33.0	9.0[b]		
LE	30.0			

[a] Two ties.
[b] Four ties.
[c] Significant at .01 level.

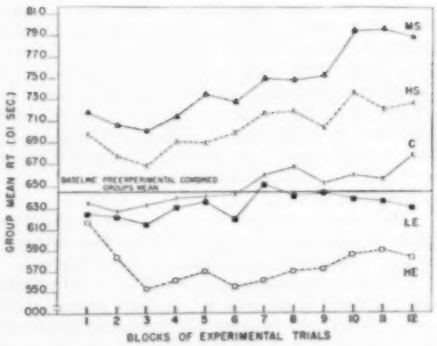

FIG. 1. Mean of the medians of each group for each of the 12 blocks of the experimental condition. (Legend: MS = Medium Stimulation, HS = High Stimulation, LE = Low Escape, HE = High Escape, C = Control.)

The mean of the medians of each of the 12 blocks of the experimental condition for a group was computed and plotted (Figure 1). To determine if the changes over trials apparent in Figure 1 were statistically significant a Trials × Subject analysis of variance (Lindquist, 1953) was performed for each group. For the analysis each of the 12 blocks was considered a treatment. The obtained F value for each group with 11 and 154 df is as follows: MS: $F = 7.672$, HS: $F = 4.106$, LE: $F = 1.997$, HE: $F = 5.236$, C: $F = 4.812$. For all but the LE group, the obtained F values were significant at $p < .001$; for the LE group the F value was significant at $p < .10$.

Post-experimental Performance

Because heterogeneity of form also obtained for groups in the post-experimental scores, a Friedman two-way analysis of variance was used to test for differences among the post-experimental Overall Means (see Table 1). A χ^2_r of 29.27 was obtained, significant at $p < .01$. Wilcoxon matched-pairs signed-ranks tests were then performed to test differences between individual pairs of post-experimental Overall Means. The results

are given in Table 2. Briefly, the MS and LE groups exhibited slower RTs than the other groups, with the LE group significantly different from the MS group. To determine if there was a significant difference between the last of the four (the worst) blocks of the experimental condition and the Overall Mean of the post-experimental condition of the MS group, a t test for differences between matched pairs was calculated and found to be significant at $p < .01$ ($t = 4.664$), indicating some improvement for the MS group during the post-experimental condition, compared to its worst experimental performance.

DISCUSSION

The present findings support the assumption that improvement on simple psychomotor tasks by chronic schizophrenics responding under aversive conditions is a function of reinforcement. It must be noted that motivation must be extremely high and reinforcement must be present, i.e., arousal or motivation alone is not a sufficient agent in improving the performance of such patients; in fact, its effects are deleterious. It may then be asked on the one hand why these individuals exhibit a general inefficiency which to some degree can be ameliorated with a negative incentive, and on the other hand why their characteristic poor behavior is at its worst with the introduction of a mild extraneous stimulus.

With respect to the problem of motivation and reinforcement, perhaps Mednick's (1958) concept of the conditioning of irrelevant thought processes in schizophrenics, coupled to Shakow's (1950) notion of their faulty set, in part explains their difficulty. Mednick suggests that the conditioning of a preoccupied and idiosyncratic condition in the schizophrenic is his means of avoiding anxiety. Instituting a steady state through the use of these responses to allay anxiety, the chronic schizophrenic has reinforced their use. Employing such safeguards leads to inadequate appraisal of and inadequate action toward incoming stimuli. The introduction of an aversive stimulus, as in the present study, can be coped with in a relatively simple manner by escaping, but is probably managed by the chronic schizophrenic in two ways. The first is the appropriate response of swift escape; the second, the inappropriate stepping up of irrelevant covert and overt Rs to reduce the attendant anxiety evoked by the subject's startle reaction to the stimulus. The latter vitiates in some measure the set to make the appropriate escape response as swiftly as possible; this failure is termed response disorganization. Thus Heath (1956) found that the degree of response disorganization was positively correlated with anxiety thresholds.

Faulty set, a symptom of schizophrenia and other neuropsychiatric disorders, and in Pribram's (1960) terms the "confusion of intentions" as to how sequences of actions are to be executed, may in part be independent of anxiety. Initially it may be an outcome of impaired servomechanisms due to physiological or learning factors or both in which there is either an insensitivity to error or a registering of errors irrespective of the situation (Pribram, 1960). The original faulty set for schizophrenics is enhanced by the reinforcement of irrelevant Rs used to assuage the anxiety which accompanies the failure to successfully cope with the situation. These severe measures needed to restrain the panic produced by constant failures aggravate the initial difficulty and result in the ultimate extremely impaired set of the chronic schizophrenic. Again, Heath's (1956) findings are appropriate.

With respect to extraneous stimuli, as in the MS and HS treatments, their introduction results in a conspicuous decline in performance. For the HS condition the schizophrenic finds no explicit or specific correction procedures which may be used in the stimulation conditions, so that the effects of the high intensity stimulus are resisted by his characteristically inappropriate behavior under trying situations, viz., thought irrelevancies. The greater slowing down induced by the MS treatment is difficult to explain. Perhaps the subject is not aware of its insidiousness and consequently suffers. This easy falling off in performance again illustrates the inadequate compensatory mechanisms of the chronic schizophrenic in coping with even the slightest perturbations in his equilibrium, and may indicate that only the grosser deviations are registered by him.

Considering that each experimental treatment lasted for 15 minutes without any interspersed rest periods, the most obvious factor which explains the continual slowing down of the two stimulation groups is fatigue. For the HE group however it might be fatigue or adaptation to the stimulus, which apparently occurred since most subjects reported that the noise did not sound so obnoxious after a while. It is difficult to determine why the LE group did not slow down at any time. It is possible that with neutral stimuli information which accumulates from several sources (in this case the visual and auditory modalities) results for chronic schizophrenics in a low level steady output. Such a notion corresponds to the highly structured and concrete environment in which they perform best. Illustrative of this idea is Leventhal's (1959) study in which better performance for chronic schizophrenics obtained when subjects were verbally rewarded and punished, while for normals either sufficed as well.

Clearly, no improvement was maintained beyond the HE experimental condition, apparently indicating that high motivation or drive and reinforcement of a concrete, specific nature must be maintained to improve the performance of chronic schizophrenics. It is also possible that the tranquilizers may have suppressed the carry-over of the improved performance, for Mitchell and Zax (1959) found that patients given chlorpromazine were less conditionable than controls. The poorer performance found in the LE and MS groups cannot be explained by a factor common to both treatments. Perhaps these findings might not reappear if the study were replicated.

SUMMARY

The present study dealt with the issue of whether the improvement in the performance of chronic schizophrenics under intense stimulation characterized as aversive is a function of reinforcement from the cessation of the stimulus or solely an increase in the level of arousal. Eighty chronic schizophrenics performed a visual two-choice RT task. Subjects were equated on RT scores in the pre-experimental condition. During the experimental condition, subjects responded under one of the following treatments: Medium or High Stimulation, in which an auditory stimulus was presented constantly throughout the experimental condition; Low or High Escape, in which an auditory stimulus was presented from which the subject could escape; and Control, which had no such collateral stimulus. The first and third treatments used a moderately intense auditory stimulus with the task; the second and fourth, a highly intense auditory stimulus. A post-experimental condition, which was the same for all groups and like the pre-experimental condition, followed.

During the experimental treatments, the Control and Low Escape groups performed the same; the High Escape group performed better than the Control group; and the two stimulation groups did worse than the Control. For the post-experimental condition the Medium Stimulation and Low Escape groups performed worse than the others which were not different from each other.

It was concluded that the improvement in the psychomotor performance of chronic schizophrenics presented with an intense stimulus characterized as aversive was a function of reinforcement rather than solely of an increase in the level of arousal.

REFERENCES

BRUCK, M. A. Electroencephalographic changes in chronic schizophrenics under chemotherapy. *Amer. J. Psychiat.*, 1958, **114**, 945.

COHEN, B. C. Motivation and performance in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 186–190.

HEATH, D. H. Individual anxiety thresholds and their effect on intellectual performance. *J. abnorm. soc. Psychol.*, 1956, **52**, 403–408.

HEBB, D. O. Drives and the c.n.s. (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–254.

HULL, C. L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951.

LANG, P. J. The visual disjunctive reaction time of normal and chronic schizophrenic subjects under varied presentations of an aversive sound stimulus. Unpublished doctoral dissertation, University of Buffalo, 1958.

LANG, P. J. The effect of aversive stimuli on reaction time in schizophrenia. *J. abnorm. soc. Psychol.*, 1959, **59**, 263–268.

LEVENTHAL, A. M. The effects of diagnostic category and reinforcer on learning without awareness. *J. abnorm. soc. Psychol.*, 1959, **59**, 162–166.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, **64**, 276–287.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

MALMO, R. B., SHAGASS, C., BELANGER, J. D., & SMITH, A. A. Motor control in psychiatric patients under experimental stress. *J. abnorm. soc. Psychol.*, 1951, **46**, 539–547.

MALMO, R. B., SHAGASS, C., & SMITH, A. A. Responsiveness in chronic schizophrenics. *J. Pers.*, 1951, **19**, 359–375.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychol. Bull.*, 1958, **55**, 316–327.

MITCHELL, L., & ZAX, M. The effects of chlorpromazine on GSR conditioning. *J. abnorm. soc. Psychol.*, 1959, **59**, 246–249.

MONROE, R. R., HEATH, R. G., MICKLE, W. A., & MILLER, W. A. A comparison of cortical and subcortical brain waves in normal, barbiturate, reserpine and chlorpromazine sleep. *Ann. N. Y. Acad. Sci.*, 1955, **61**, 56–71.

PRIBRAM, K. H. A review of theory in physiological psychology. *Annu. Rev. Psychol.*, 1960, **11**, 1–40.

ROSENBAUM, G., GRISELL, J. L., & MACKAVEY, W. R. Effects of biological and social motivation on schizophrenic reaction time. *J. abnorm. soc. Psychol.*, 1957, **54**, 364–368.

SHAKOW, D. Some psychological features of schizophrenia. In M. L. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950. Pp. 383–390.

WHITEHEAD, W. A., & THUNE, L. E. The effects of chlorpromazine on learning in chronic psychotics. *J. consult. Psychol.*, 1958, **22**, 379–383.

(Received June 1, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 112–120

ATTITUDE CHANGE AND MENTAL HOSPITAL EXPERIENCE

JACK M. HICKS AND FRED E. SPANER[1]

Northwestern University *Veterans Administration Hospital, Downey, Illinois*

It is the thesis of this investigation that favorable attitudes toward the mentally ill may develop as a function of intimate exposure to the mental hospital environment. Specifically, attitudes of hospital employees will shift over time in a direction more favorable to the patient. Empirical support of this notion is suggested by Bovard (1951), who found that as frequency of interaction between persons increases, the degree of liking for one another increases. His study, however, involved college classes and may not be generalizable to the mental hospital.

Impressionistic evidence of the positive effects of mental hospital experience is provided by extensive work with student nurses (Perlman & Barrell, 1958). These observers report that at the beginning of psychiatric training student nurses see patients as "maniacal" and tend to emphasize the differences between patients and themselves. But, as their training progresses, "they begin to maximize the similarities between themselves and patients, and see the differences as predominantly differences in degree and not kind." Some experimental corroboration of these observations is offered by Altrocchi and Eisdorfer (1960).

Dissenting evidence of the positive effects of the hospital environment upon attitudes is suggested by Middleton (1953) who found greater prejudice toward mental patients among older, more experienced mental hospital employees than among younger, less experienced employees. However, this finding is difficult to interpret as evidence of attitude change due to the absence of pretest information.

In view of the paucity of systematic research and ambiguity of findings in this area, the present investigation was undertaken.

[1] The authors would like to express their indebtedness to Donald T. Campbell who critically read this paper and made many valuable suggestions.

Experiment I

This study was an initial step toward further exploration of the effects of mental hospital experience upon attitudes toward the mentally ill. Attitude modification was measured among student nurses exposed to 12 weeks of psychiatric nursing training at the Veterans Administration Hospital, Downey, Illinois. Attitudes elicited immediately prior to training were compared to attitudes immediately following training.

The following hypothesis was tested:

There will be a favorable shift in attitude toward mental patients over a 12-week period of psychiatric nursing training.

Favorable attitude shift was defined as a change in response to attitude statements in the direction of judges' opinions of the most desirable attitudes for student nurses to have.

Psychiatric nursing training consisted of two main subdivisions: classroom instruction, and ward experience. Classroom instruction involved 10 hours per week devoted to exploration of nursing care problems. Ward experience entailed 6 weeks on each of two clinical areas—the male section and the female section—of the Acute and Intensive Treatment Service.

Method

Subjects. Seventy-eight student nurses, sent to Downey from several schools of nursing in the Chicago area as part of their regular nursing training schedule, were used. They were there in groups of 48 and 30, 3 months apart. All had received an average of about 2 years' previous "Nonpsychiatric" nursing training.

Materials. A five-point Likert questionnaire designed to elict attitudes associated with mental patients was used. The scale contained 66 statements some of which were borrowed verbatim or modified from Gilbert and Levinson's (1956) CMI scale (Items 5, 7, 11, 15, 18, 19, 25, 31, and 38) and Middleton's (1953) "Prejudice Test" (Items 1, 2, 5, 10, 14, 17, 22, 23, 26, 34, 36, and 37). The remaining items were original with this study.

Design. The design for this study, presented in Table 1, was a "recurrent institutional cycles design"

developed by Campbell and Stanley (1962). It is appropriate to those situations in which an institutional process (in this case, psychiatric nursing training) is presented continually to new groups of respondents. Group A was a posttest only aggregate whose only exposure to the questionnaire was on the final day of their 12-week psychiatric training period. Group B was a pretest-posttest aggregate. These respondents received the questionnaire on the first morning of their arrival into the psychiatric setting, prior to any formal orientation or exposure to patients (Group B_1). They also received an identical form of the questionnaire on the final day of their training period (Group B_2).

Procedure. First exposure: For the pretest and posttest only conditions, the experimenter was introduced to the group by a nursing education instructor. The experimenter then asked their cooperation in participating in a survey of attitudes and opinions about mental illness. Confidentiality of results was assured. Group B_1 were asked to sign their names on the questionnaire. Group A were asked not to record names.

Second exposure: Group B_2 respondents, having already been exposed to the test situation, were not given the introduction and full instructions. The group was simply asked to participate in filling out another questionnaire concerning attitudes and opinions about mental illness. They were reminded to put their names on the questionnaires. The same test room and same experimenter were used for all conditions.

Composite scoring key: A total attitude score was derived for each respondent by the method of summated ratings. The directionality of each item was determined by asking five nursing education instructors to respond to each statement in accordance with their professional opinions as to the most desirable attitude for a student nurse to have. These ratings produced at least 80% interjudge agreement on 38 of the total 66 items. Item response categories were scored from zero through four, depending on the directionality of the statement.

Results and Discussion

The major statistical tests of the hypotheses shown in Table 1 were t tests of difference in mean composite scores between (*a*) Group B_1 (pretest) and Group B_2 (posttest), and (*b*) Group B_1 and Group A (posttest only).

The mean difference between B_1 and B_2 was 15.25. A direct-difference t test between these two means was highly significant and in the predicted direction ($p < .0005$). This finding strongly supported the hypothesis. The mean difference between B_1 and A was 4.5, reaching significance in the same direction as the B_1–B_2 difference ($p < .05$). Hence, support of the hypothesis was provided by both comparisons.

TABLE 1

COMPARISONS OF COMPOSITE SCORE MEANS

Comparison	*N*	*M*	*t*
B_1 (Pretest)	48	103.1	11.55***
B_2 (Posttest)	48	118.3	
A (Posttest only)	30	107.6	1.96*
B_1	48	103.1	
A	30	107.6	4.67**
B_2	48	118.3	

* Significant at .05 level, one-tailed.
** Significant at .001 level.
*** Significant at .0005 level, one-tailed.

However, these results became less clear-cut in consideration of an unexpected divergence between the two posttest means. A t test between A and B_2 produced a convincingly significant difference ($p < .001$), with B_2 higher than A. This finding seemed noteworthy despite the fact that a t test did not show the unpredicted A–B_2 difference to be significantly greater than the predicted A–B_1 difference ($p > .05$). There appeared to be several plausible explanations. One possibility stemmed directly from the fact that B_2 received a pretest, whereas A did not. That is, the pretest may have influenced performance on the subsequent testing. According to Campbell (1957), a pretest may increase or decrease sensitivity to the experimental variable. Thus, there may have been an interaction between the attitude pretest and psychiatric experience.

Another possible cause of the A–B_2 discrepancy was a respondent selection bias. Group B may have had a higher mean attitude than Group A at the outset of psychiatric training. This consideration became particularly important since the nature of the field situation prevented random preselection by the experimenter. Although psychiatric nursing trainee assignments were made unsystematically, and were not expected to vary from group to group, this assumption could only be made on a priori grounds at this time.

A third source of confoundment pertained to names being asked of Group B_2 but not Group A. This fact may have "freed" Group A respondents to more readily express less favorable attitudes actually felt. Group B_2, on the other hand, because of the presence of their names, may have been more inclined

to present themselves in a favorable light, giving answers which they knew the staff thought desirable.

A fourth possibility concerned the sheer fact of taking a pretest sufficing to produce systematic changes on the posttest. This phenomenon, differs from test-treatment interaction in that it refers to the main effects of testing per se.

Finally, it is possible that "history" effects of a powerful extraneous influence unknown to the experimenter occurred prior to B_2, but not prior to A. An example would be a mental health publicity campaign occurring contemporaneously with the Group A training period. A related possibility is that psychiatric training itself differed in some important way for Groups A and B. It is certainly possible to imagine that some change in hospital personnel, goals, or other conditions, may have affected the nursing education program.

EXPERIMENT II

This investigation was an attempt to (*a*) corroborate and refine the basic findings of Experiment I, and (*b*) clear up ambiguities in the results of Experiment I. It will be recalled that a small but significant difference in the predicted direction was found between a posttest only and an independent pretest group, but that a highly significant difference between the two posttests created problems of interpretation. Basically, the same quasi-experimental design was used (institutional cycles design), but with additional controls and a modified measuring instrument. This design also resembles Campbell and Stanley's (1962) "nonequivalent control groups" design in that random preselection was not possible.

The primary advantage of Experiment II was in the introduction of control groups. This addition made two important contributions: First, it provided for the testing of a hypothesis more precisely related to the effects of psychiatric training. The hypothesis for Experiment I was limited to predicting "absolute" attitude shift over 12 weeks of psychiatric training. The study was not designed to differentiate between change as a function of psychiatric training, and change produced by other variables. As much change may conceivably have occurred with no psychiatric training at all. Equipped with control groups, Experiment II may test the more meaningful hypothesis of whether or not attitude change will be relatively greater as a function of psychiatric training than some other treatment. Hypothesis I for this investigation is:

There will be greater favorable shift in attitudes toward mental patients over a 12-week period of psychiatric nursing training than over an equivalent period of nonpsychiatric nursing training. (Attitude change was defined as in Experiment I.)

A second advantage of control groups permitted light to be cast on factors related to the unpredicted mean difference between the posttests of Experiment I. In particular, it was possible to test for pretest "sensitization" to psychiatric experience. For Experiment I, there were pretested and unpretested groups for the experimental treatment only. For Experiment II, there were pretested and unpretested groups for both experimental and control conditions. Due to the suggestive finding of Experiment I, the following hypothesis may be stated:

There will be a significant interaction between pretest attitude and psychiatric experience. This interaction should have a favorable rather than a dampening effect upon posttest attitudes.

An alternative possibility which can be tested as part of the analysis of the above hypothesis concerns the main effects of testing. In this consideration, pretested respondents would have significantly more favorable attitudes than unpretested respondents to a comparable degree on both experimental and control conditions.

Experiment II was also designed to test for recruitment differences among groups of respondents, another of the suggested confounders of Experiment I. This analysis was made by testing for nonlinearity among mean attitudes of all pretest conditions, both Experimental and Control.

It was also conjectured that the unpredicted difference between the posttests of Experiment I was due to instructions to record names on the questionnaires for one group but not the other. In order to check out this

possibility, an additional experimental group was tested with "no name" instructions.

One further test designed to clarify Experiment I was considered. The reader will recall the suggestion that the two posttest means differed due to "history" effects of extraneous influences occurring concomitantly with psychiatric training, or subtle changes in psychiatric training itself. Such confounders were controlled for in Experiment II by the inclusion of two pretest-posttest psychiatric groups differing in terms of a 3-month separation in experimental treatment. If either or both of these effects are manifest, differential attitude change between the two groups would be expected.

Method

Subjects. A total of 354 student nurses were used. Two hundred twenty-four were members of the experimental groups receiving 3 months of intensive psychiatric nursing training. The remaining 130 were control respondents who, not yet having reached the psychiatric phase of nursing training, were encountered somewhat earlier in their overall programs. There was considerable overlap in this respect, however, since entrance of students into psychiatric training was distributed fairly evenly over the second half of the 3-year registered nurses training program. The only prerequisite for psychiatric training was at least 18 months of previous general hospital training.

All Experimental subjects took their psychiatric training at the Veterans Administration Hospital, Downey, Illinois, as part of their normal nursing training schedule. They came in aggregates of approximately 50 at 3-month intervals from seven schools of nursing in the Chicago area. Control respondents were tested while undergoing pediatric, surgical, obstetrical, or other phases of nursing training at two of the seven home hospitals of the nursing schools mentioned above. Respondents were not randomly assigned by the experimenter.

Materials. The Experiment I questionnaire was used except for slight modifications in format and content. In terms of format, the six-point forced-choice Likert adaptation, employed by Cohen and Struening (1959) was adopted. This represented an elimination of the ? category used too liberally by Experiment I respondents, being replaced by two categories: not sure but probably agree, and not sure but probably disagree.

The content of the questionnaire was modified to the extent of eliminating 5 items and adding 26 new ones, making a total of 87 for the Experiment II questionnaire. The 5 eliminated items were unsatisfactory because of highly skewed response distributions. One of these items was Number 22 from Middleton's (1953) Prejudice Test. The addition of 26 items was an attempt for better representation of the attitude universe of interest. Of these items, 4 were borrowed from Middleton's Prejudice Test (Items 4, 18, 29, and 30), and 8 were taken from Cohen and Struening's (1959) Opinions about Mental Illness questionnaire (Items 8, 14, 34, 35, 39, 41, 42, and 48). The remaining 14 items were original with this study. A special feature of the revised questionnaire was the retention of 21 nonshifting items from Experiment I. These items were included as an indicator of the degree to which Experimental respondents discriminate among items in the process of demonstrating attitude change. Fourteen items which showed significant change in the same direction on both $B_1 - A$ and $B_1 - B_2$ comparisons in Experiment I were also retained.

Design. The 354 respondents made up eight conditions, five treatment conditions and three control conditions. Experimental groups are referred to in Table 2 as Groups A, B, F, G, and H. Groups C, D, and E are Control groups. The lettering, A through H, is related to the chronological order in which the different groups were encountered over a 9-month period. The Experimental groups were of two types: pretest-posttest and posttest only. Groups A and H (comparable to Group A of Experiment I) were posttest only (or unpretested) whose only exposure to the questionnaire was on the final day of their respective 12-week psychiatric training periods. Groups A and H differed in that A was tested 9 months prior to H; and Group A respondents were asked to record their names on the questionnaires, while H respondents remained anonymous. Groups B and G were Experimental pretest-posttest aggregates, comparable to Group B of Experiment I. These subjects were pretested immediately upon arrival at Downey and prior to any of the scheduled activities (Conditions B_1 and G_1). They were posttested on an identical form of the questionnaire after completion of all scheduled psychiatric activities 12 weeks later (Conditions B_2 and G_2). A 3-month separation in the span of time (12 weeks) during which training was received denotes the principal difference between these two groups.

In order to determine the longer range effects of psychiatric experience upon attitude changes, an additional Experimental group (Group F) was tested. Of the 24 individuals in this group, 14 had participated previously in Experiment I and 10 in Group B of Experiment II, discussed above. These subjects ranged from 1 to 10 months of "post-psychiatric" general hospital training, with a mean of 5.25 months.

Control groups were also of two types: pretest-posttest and pretest only. There were two pretest only groups (C and E), so-called as a convenient label for "one-shot" testing administered at times prior to psychiatric training. Minor differences between Groups C and E were in terms of time and hospital at which testing took place. The final condition to be discussed pertains to Group D, a nonpsychiatric pretest-posttest aggregate. This group,

TABLE 2

EXPERIMENTAL DESIGN

Groups[a]							
A	B	C	D	E	F	G	H
Treatment	↓	↓	↓	↓	(Treatment)	↓	↓
Posttest	Pretest				(Posttest)		
	Treatment	Pretest	Pretest		↓	Pretest	
	Posttest		↓	Pretest	Post-Posttest	Treatment	Treatment
			Posttest			Posttest	Posttest[b]

[a] In chronological order.
[b] Respondents asked not to record names.

representing the basic control for testing of Hypothesis I, received 12 weeks of general hospital training between pretest and posttest.

Procedure. The identical procedure was used as described in Experiment I. The same experimenter presided over all sessions, with one exception. A substitute examiner was used for Condition G_2. Instructions were the same. All groups were requested (on both pre- and posttest) to record their names in the upper right-hand corner of the questionnaire except Group H. Group H was asked not to record names.

As in Experiment I, the method of summated ratings was used. Directionality of the new attitude statements was established by the same five nursing education instructors used in Experiment I. Items retained from Experiment I were not "rekeyed." Eighty percent interjudge agreement was achieved on 59 of the 87 items. Items were scored from one through six for analysis purposes.

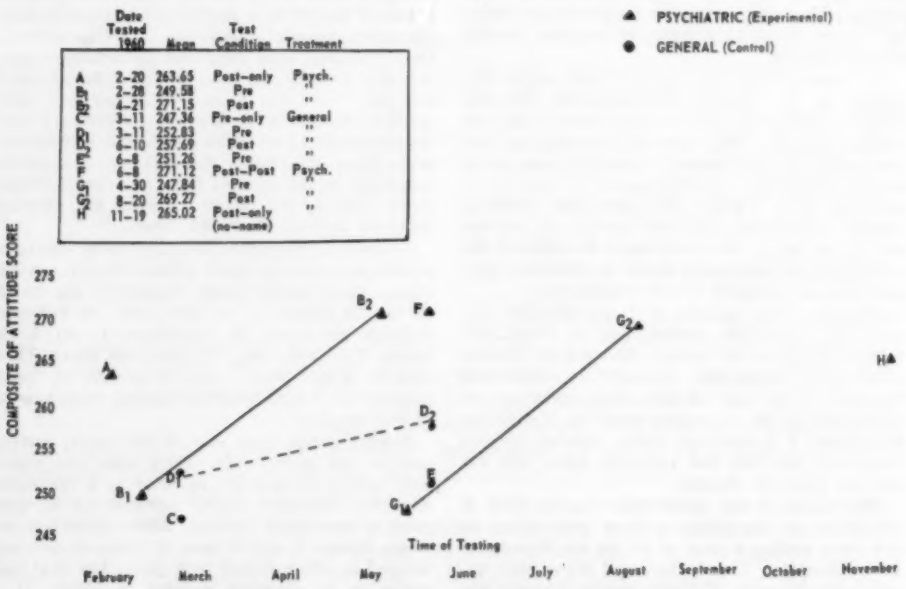

FIG. 1. Mean attitudes and attitude shifts for all groups.

Results

A post hoc measure of test-retest stability of the 59 keyed items was obtained by matching pretest-posttest scores for Control Group D. The Pearson coefficient was .82. Similar Pearson *r*'s were obtained between pretests and posttests of Experimental Groups B and G of .55 and .57, respectively.

Internal consistency of the keyed items was estimated as a by-product of average item-total correlations (Guilford, 1950, p. 494). Two independent matrices of item-total correlations were computed. Groups A, B, and G were combined, as well as Groups C, D, and E, yielding average item-total correlations of .34 and .30, respectively. Resulting internal consistency reliabilities were .88 and .86. The average item-total *r*'s were also squared to provide estimates of .11 and .09 mean item intercorrelations.

An overall picture of mean attitudes and attitude shifts is presented in Figure 1. A cursory inspection reveals both markedly higher attitude scores for psychiatric groups than Control groups overall, and sharp increments in attitude favorability over the 12-week psychiatric training period. Mean attitude increments for Experimental Groups B and G were 21.57 and 21.43, respectively, both of which are highly significant ($p <$ .0005, one-tailed), thus, strongly supporting the hypothesis of Experiment I.

TABLE 3

MULTIPLE COMPARISONS TEST OF DIFFERENTIAL ATTITUDE SHIFT

Comparison	Difference	$\overline{WSD}$
$(\bar{X}_{B_1}-\bar{X}_{B_2})-(\bar{X}_{D_1}-\bar{X}_{D_2})$	16.71	9.37*
$(\bar{X}_{G_1}-\bar{X}_{G_2})-(\bar{X}_{D_1}-\bar{X}_{D_2})$	16.57	8.81*
$(\bar{X}_{B_1}-\bar{X}_{B_2})-(\bar{X}_{G_1}-\bar{X}_{G_2})$	.14	8.81

* Significant at .01 level.

The Tukey test for multiple comparisons, reviewed by Ryan (1959), was employed for the assessment of Hypothesis I. A three-way comparison was made between Experimental Groups B and G, and Control Group D in terms of degree of favorable attitude shift.

Groups B and G exceeded D in magnitude of favorable attitude shift by mean differences of 16.71 and 16.57, respectively. Both of these increments exceeded their respective "wholly significant difference" ($\overline{WSD}$) set at the .01 level of confidence indicated in Table 3. The third comparison between Groups B and G produced no significant difference, as expected. These comparisons unequivocally supported Hypothesis I as shown graphically in Figure 2.

Pretest-treatment interaction was tested by

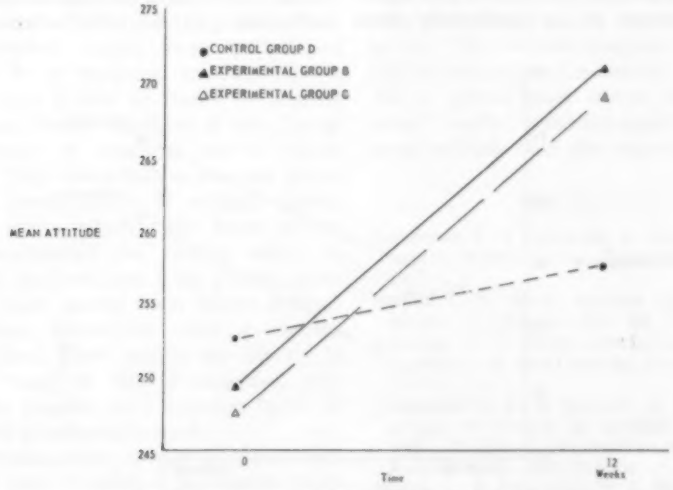

FIG. 2. Mean attitude shift over 12-week period.

TABLE 4

ANALYSIS OF VARIANCE OF MEAN ATTITUDE DIFFERENCES AS A FUNCTION OF TESTING AND TREATMENT

Source	*df*	*MS*	*F*
Pretest	1	3932.5	11.00*
Treatment	1	14,543.3	40.68*
Interaction	1	140.2	.39
Error (w)	326	357.5	

* Significant at .001 level.

a fourfold analysis of variance for unequal and disproportionate cell frequencies. Experimental versus Control groups made up the row main effects, whereas pretested versus unpretested groups were represented in the columns. As noted in Table 4, and also graphically in Figure 3, there was no significant interaction between pretesting and psychiatric nursing experience ($F < 1$). Therefore, Hypothesis II was rejected in that there was no evidence of pretest sensitization to the experimental variable.

Table 4 does present, however, strong evidence of the main effects of testing. Mean posttest attitudes of pretested respondents were found to be significantly higher than corresponding mean attitudes of posttest only subjects ($p < .001$).

The significant main effects of the experimental variable are also shown in Table 4. Overall mean posttest attitudes of psychiatric nurses are shown to be significantly more favorable than posttest attitude scores of the nonpsychiatric groups. This analysis does not reflect attitude *shift*, however, and does not bear directly on Hypothesis I.

TABLE 5

ANALYSIS OF VARIANCE OF RECRUITMENT DIFFERENCES

Source	*df*	*MS*	*F*
Between groups	4	252.5	<1
Error	234	338.9	

Results of a one-way analysis of variance are presented in Table 5, testing for nonlinearity among group mean pretest attitudes. The between-groups variance was less than the within-groups variance, hence rejecting the nonlinearity hypothesis ($F < 1$) regarding recruitment nonequivalence.

The hypothesized difference between "names" and "no names" was tested by comparing the attitude means of posttest only Experimental Groups A and H. The means for the two groups were 263.65 and 265, respectively. An uncorrelated t was < 1, failing to support the hypothesis.

The effects of "history" and/or undetected changes in the experimental variable were tested by comparing the mean shifts of two psychiatric groups, B and G. A test for this hypothesis was conveniently provided by Table 3. As previously mentioned, the mean difference (.14) in shifts between Groups B

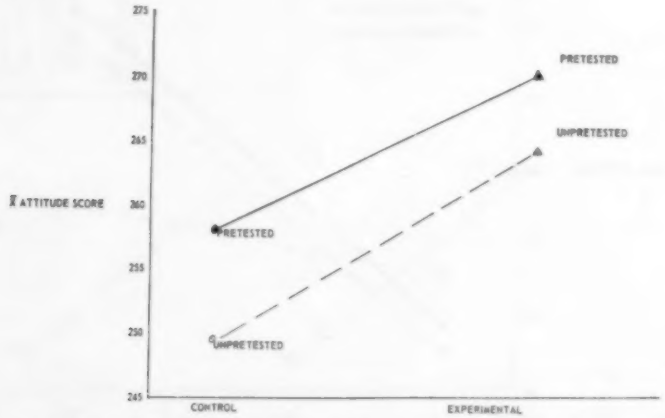

FIG. 3. Conjunctive effects of testing and treatment on posttest scores.

and G is far below the $\overline{WSD}$ of 8.81. Thus, no history effects or changes in psychiatric training appear evident.

The testing of 24 "post-posttest" respondents (Group F) made it possible to estimate the stability of attitude change. Since most of these respondents were previous members of psychiatric pretest-posttest conditions, they were compared accordingly with pretest-posttest Groups B and G. The post-posttest mean attitude of Group H was 271.12, a level remarkably comparable to a posttest mean of 271.15 for Group B and 269.27 for Group G. The durability of attitude change, at least for a period of several months, seems clearly demonstrated.

A comparison of degree of shifts was also made between shifting and nonshifting items of Experiment I. The criterion of item shift on Experiment II was a significant mean difference on the B_1–B_2 comparison. A Yates-corrected χ^2 was a nonsignificant 1.73 ($p > .05$). Hence, items which shifted on Experiment I did not shift to a significantly greater extent on Experiment II than items which did not shift on Experiment I.

DISCUSSION

Of the several additional subsidiary hypotheses concerning the unpredicted difference between the posttests of Experiment I, only that which concerned pretest main effects was supported. It appears, therefore, that the significantly higher mean attitude of Condition B as compared to Condition A of Experiment I may be traced to transfer effects from Pretest Condition B with Group A not being so benefited by a pretest "booster." This interpretation does not follow from two investigations of attitude change by Lana (1959a, 1959b), who found neither pretest sensitization nor testing effects. A third study by Lana and King (1960), however, did show testing main effects without test-treatment interaction when a learning task was used. These results are directly in line with those of this investigation, suggesting the presence of a learning factor in the attitude questionnaire used.

In conclusion, then, it may be stated with confidence that 12 weeks of psychiatric training was demonstrated to be unequivocally effective in producing attitude change toward mental illness over and above the noise level of pretest transfer effects. Furthermore, the absence of troublesome interactions strengthens the generalizability of these findings beyond this particular experimental arrangement.

SUMMARY

Two quasi-experiments were performed as part of an on-going project to investigate the effects of overall mental hospital experience upon modification of attitudes toward the mentally ill. Four hundred thirty-two student nurses served as respondents over all. Attitudes were measured by the Likert technique before and after 12 weeks of psychiatric nursing training. In Experiment I, attitudes on both posttests were observed to be significantly higher in a favorable direction than a pretest. A complication arose, however, with respect to an unpredicted significant difference between the posttests. Experiment II was basically a refinement of Experiment I. Control groups were added to the design, the questionnaire was revised, and a larger N used. Control respondents were student nurses in general hospitals not as yet having reached the psychiatric phase of their training. The basic hypothesis that psychiatric subjects would make a significantly greater favorable shift in attitudes over a 12-week period than nonpsychiatric subjects was strongly supported. The evidence suggests that the posttest discrepancy in Experiment I was probably due to pretest main effects, with the "pretested" posttest condition manifesting a higher mean attitude than the unpretested group.

REFERENCES

ALTROCCHI, J., & EISDORFER, C. Changes in attiudes toward mental illness. Unpublished manuscript, 1960.

BOVARD, E. W. Group structure and perception. *J. abnorm. soc. Psychol.*, 1951, **46**, 398–405.

CAMPBELL, D. T. Factors relevant to the validity of experiments in social settings. *Psychol. Bull.*, 1957, **54**, 297–312.

CAMPBELL, D. T., & STANLEY, J. C. Experimental designs for research in teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand-McNally, 1962, in press.

COHEN, J., & STRUENING, E. L. Factors underlying opinions about mental illness in the personnel of a

large mental hospital. Paper read at American Psychological Association, Cincinnati, 1959.

GILBERT, DORIS, & LEVINSON, D. J. Ideology, personality, and institutional policy in the mental hospital. *J. abnorm. soc. Psychol.*, 1956, **53**, 263–271.

GUILFORD, J. P. *Fundamental statistics in psychology and education.* (2nd ed.) New York: McGraw-Hill, 1950.

LANA, R. E. A further investigation of the pretest-treatment interaction effect. *J. appl. Psychol.*, 1959, **43**, 421–422. (a)

LANA, R. E. Pretest-treatment interaction effects in attitudinal studies. *Psychol. Bull.*, 1959, **56**, 293–300. (b)

LANA R. E., & KING, D. J. Learning factors and determiners of pretest sensitization. *J. appl. Psychol.*, 1960, **44**, 189–191.

MIDDLETON, J. Prejudices and opinions of mental hospital employees regarding mental illness. *Amer. J. Psychiat.*, 1953, **10**, 133–138.

PERLMAN, M., & BARRELL, L. M. Teaching and developing nurse-patient relationships in a psychiatric setting. *Psychiat. quart. Suppl.*, 1958, **31**, 1–13.

RYAN, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, **56**, 26–47.

(Received June 3, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 121–128

EPINEPHRINE, CHLORPROMAZINE, AND AMUSEMENT[1]

STANLEY SCHACHTER AND LADD WHEELER

Columbia University *University of Minnesota*

In their study of cognitive and physiological determinants of emotional states, Schachter and Singer (1962) have demonstrated that cognitive processes play a major role in the development of emotional states. Given a common state of physiological arousal, subjects can be readily induced into states of euphoria or of anger by means of cognitive manipulations. To what extent the state of physiological arousal is a necessary component of an emotional experience is not, however, completely clear in that study.

The technique employed by Schachter and Singer (1962) to produce a state of physiological arousal was simply the injection of the sympathomimetic amine, epinephrine. With slight exceptions, this agent provokes a pattern of physiological activation which is a virtual replica of the state produced by active discharge of the sympathetic nervous system. In experimental situations designed to make subjects euphoric, those subjects who received injections of epinephrine were, on a variety of indices, somewhat more euphoric than subjects who received a placebo injection. Similarly, in situations designed to make subjects angry and irritated, those who received epinephrine were somewhat angrier than subjects who received placebo. In both sets of conditions, however, these differences between epinephrine and placebo subjects were significant, at best, at borderline levels of statistical significance.

Assuming, for the moment, that physiological arousal is a necessary component of emotional states, one of the factors that might account for this failure to find larger differences between epinephrine and placebo subjects seems reasonably apparent. The experimental situations employed were fairly effective. The injection of placebo does not, of course, prevent the subject from self-arousal of the sympathetic system, and indeed there is considerable evidence (Woodworth & Schlosberg, 1958) that the arousal of an emotional state is accompanied by general excitation of the sympathetic nervous system.

A test of the proposition at stake, then, would require comparison of subjects who have received injections of epinephrine with subjects who, to some extent, are rendered incapable of self-activation of the sympathetic nervous system. Thanks to a class of drugs known generally as autonomic blocking agents, such blockade is, to some degree, possible. If the proposition that a state of sympathetic discharge is a necessary component of an emotional experience is correct, it should be anticipated that whatever emotional state is experimentally manipulated, it should be most intensely experienced by subjects who have received epinephrine, next by placebo subjects, and least of all by subjects who have received injections of an autonomic blocking agent.

Procedure

In order to conceal the purposes of the study and the nature of the injection, the experiment was cast in the framework of a study of the effects of vitamins on vision. As soon as a subject arrived, he was taken to a private room and told by the experimenter:

> I've asked you to come today to take part in an experiment concerning the effects of vitamins on the visual processes. We know a great deal about vision, but only night vision has been studied in relation to nutrition. Our experiment is concerned with the effects of suproxin on vision. Suproxin is a high concentrate vitamin C derivative. If you agree to take part in the experiment, we will give you an injection of suproxin and then subject your retina to about 15 minutes of

[1] This experiment is part of a program of research on cognitive and physiological determinants of emotional state which is being conducted at the Department of Social Psychology at Columbia University under PHS Research Grant M-2584 from the National Institute of Mental Health, United States Public Health Service. This experiment was conducted at the Laboratory for Research in Social Relations at the University of Minnesota. We wish to thank Abelardo Mena who was the physician in this study.

continuous black and white stimulation. This is simpler than it sounds; we'll just have you watch a black and white movie. After the movie, we'll give you a series of visual tests.

The injection itself is harmless and will be administered by our staff doctor. It may sting a little at first, as most injections do, but after this you will feel nothing and will have no side effects. We know that some people dislike getting injections, and if you take part in the experiment, we want it to be your own decision. Would you like to? [All subjects agreed to take part.]

This much said, the experimenter gave the subject a test of visual acuity and of color vision, took the subject's pulse and left the room. Shortly thereafter, the doctor arrived, gave the subject a quick ophthalmoscopic examination, then gave him an injection and informed him that the experimenter would be back for him shortly "in order to take you and some other subjects who have also received shots of suproxin into the projection room."

Injections

There were three forms of suproxin administered—epinephrine, placebo, and chlorpromazine.

1. Epinephrine: Subjects in this condition received a subcutaneous injection of ½ cubic centimeter of a 1:1000 solution of Winthrop Laboratory's Suprarenin.

2. Placebo: Subjects in this condition received a subcutaneous injection of ½ cubic centimeter of saline solution.

3. Chlorpromazine: Subjects in this condition received an intramuscular injection of a solution consisting of 1 cubic centimeter (25 milligrams) of Smith, Kline, and French Thorazine and 1 cubic centimeter of saline solution.

The choice of chlorpromazine as a blocking agent was dictated by considerations of safety, ease of administration, and known duration of effect. Ideally, one would have wished for a blocking agent whose mechanism and effect was precisely and solely the reverse of that of epinephrine—a peripherally acting agent which would prevent the excitation of sympathetically innervated structures. Though it is certainly possible to approach this ideal more closely with agents other than chlorpromazine, such drugs tend to be dangerous, or difficult to administer, or of short duration.

Chlorpromazine is known to act as a sympathetic depressant. It has a moderate hypotensive effect, with a slight compensatory increase in heart rate. It has mild adrenergic blocking activity for it reverses the pressor effects of small doses of epinephrine and depresses responses of the nictitating membrane to preganglionic stimulation. Killam (1959) summarizes what is known and supposed about the mechanism of action of chlorpromazine as follows:

> Autonomic effects in general may be attributed to a mild peripheral adrenergic blocking activity and probably to central depression of sympathetic centers, possibly in the hypothalamus (p. 27).

Popularly, of course, the compound is known as a "tranquilizer."

It is known that chlorpromazine has effects other than the sympatholytic effect of interest to us. For purposes of experimental purity this is unfortunate but inevitable in this sort of research. It is clear, however, that the three conditions do differ in the degree of manipulated sympathetic activation.

Subjects

Subjects were male college students taking classes in introductory psychology at the University of Minnesota. Some 90% of the students in these classes volunteer for a subject pool, for they receive two extra points on their final exam for every hour that they serve as experimental subjects. The records of all potential subjects were cleared with the Student Health Service in order to insure that no harmful effects would result from injections of either epinephrine or chlorpromazine.

Each experimental group was made up of three subjects—one from each of the injection conditions. Their appointments were staggered slightly so as to insure sufficient time for the particular drug to be absorbed. Thus, the chlorpromazine subject received his injection about 15 minutes before the movie began. Pretests had revealed that, with this dosage and mode of administration, about this time interval was required for the onset of sympathetic effects. Placebo subjects were injected 5–10 minutes before onset of the movie. Epinephrine subjects were injected immediately before the movie so that at most 3–4 minutes went by between the time they were injected and the beginning of the film. Pretests had shown that the effects of epinephrine began within 3–5 minutes of injection. It was, of course, basic to the experimental design that these effects begin only after the movie had started.

Film

Rather than the more complicated devices employed in the Schachter and Singer (1962) experiment, an emotion inducing film was used as a means of manipulating the cognitive component of emotional states. In deciding on the type of film, two extremes seemed possible—a horror, fright, or anxiety provoking film or a comic, amusement provoking film. Since it is a common stereotype that adrenalin makes one nervous and that the tranquilizer, chlorpromazine, makes one tranquil and mildly euphoric, the predicted pattern of results with a horror film would be subject to alternative interpretation. It was deliberately decided, then, to use a comedy. If our hypothesis is correct, it should be anticipated that epinephrine subjects would find the film somewhat funnier than placebo subjects who, in turn, would be more amused than chlorpromazine subjects.

The film chosen was a 14-minute 40-second

excerpt from a Jack Carson movie called *The Good Humor Man*. This excerpt is a self-contained, comprehensible episode involving a slapstick chase scene.

The projection room was deliberately arranged so that the subjects could neither see nor hear one another. Facing the screen were three theatre-type seats separated from one another by large, heavy partitions. In a further attempt to maintain the independence of the subjects, the sound volume of the projector was turned up so as to mask any sounds made by the subjects.

Measurement

Observation. During the showing of the movie an observer, who had been introduced as an assistant who would help administer the visual tests, systematically scanned the subjects and recorded their reactions to the film. He observed each subject once every 10 seconds, so that over the course of the film 88 units of each subject's behavior were categorized. The observer simply recorded each subject's reaction to the film according to the following scheme.

1. Neutral: Straight-faced watching of film with no indication of amusement.
2. Smile.
3. Grin: A smile with teeth showing.
4. Laugh: A smile or grin accompanied by bodily movements usually associated with laughter, e.g., shaking shoulders, moving head, etc.
5. Big laugh: Belly laugh—a laugh accompanied by violent body movement such as doubling up, throwing up hands, etc.

In a minute by minute comparison, two independent observers agreed in their categorization of 90% of the 528 units recorded in six different reliability trials. Lumping Categories 2 through 5 together, the two observers agreed on 93% of the units jointly recorded.

The observer, of course, never knew which subject had received which injection.

Evaluation of the film. The moment the movie ended the lights were turned on and the experimenter proceeded:

> Before beginning the visual tests, we want your eyes to recover somewhat from the constant stimulation they've just received. The rate of neuro-limnal recovery under conditions of perfectly normal lighting and coloring is of major interest to us. The recovery will have begun in about 12 minutes, and after that time, Dr. Mena will give you the more precise visual examination.
>
> In the meantime, I'd like to ask your help. As I told you, we need about 15 minutes of retinal stimulation, for which purpose we use a movie. Obviously, it doesn't matter at all to us which movie we use, so long as it is black and white. We can use one movie just as easily as another, but we do want to use a film that you like. I'm sure that you can see the necessity of using a film which our subjects will like. Of course, the only way to find out if you like it is to ask you. We're just beginning this experiment and will have many more subjects like you. Since you are one of the first groups, it will be a big help if you will give us your personal reactions to the film. If you like it, we'll keep it and if you don't like it, we can just as easily get another. If you'll use these mimeographed questionnaires, it will make it easier for us.

The experim.nter then handed out a questionnaire whose chief items, for present purposes, were the following:

1. How funny did you find this film?

Extremely dull (1)	Very dull (2)	Somewhat dull (3)	Mildly funny (4)	Very funny (5)	Extremely funny (6)

2. All in all, how much did you enjoy this film?

Disliked it intensely (1)	Disliked it very much (2)	Disliked it a little (3)	Enjoyed it a little (4)	Enjoyed it very much (5)	Enjoyed it enormously (6)

The figures in brackets represent the values used in computing the means presented in later tables.

3. Would you recommend that we should show this particular film to our future subjects?

(3) —— strongly recommend keeping this film.
(2) —— moderately recommend keeping this film.
(1) —— recommend you get another film.

Physical condition. In order to check on whether or not the drugs were having the desired effect on the subject's internal state, after the subjects had evaluated the film, the experimenter continued with the following spiel:

> Before we begin the eye tests, we need a bit more information about you. Earlier studies on the visual processes have shown that a person's physical and emotional states influence the visual process. Because of this it is necessary to know how you feel physically and emotionally at this time. We know, for example, that certain states such as hunger, or fatigue, or boredom, do have a noticeable effect on these processes. Naturally, we have

to know these things about you in order to interpret the results we will obtain from each of you, and the only way we can find out such things is to ask you.

A bit more on this line and the experimenter then handed round a questionnaire whose chief items were the following:

A. For evaluation of the effects of epinephrine:

1. Have you experienced any palpitation (consciousness of your own heartbeat) during the last half hour or so?

Not at all	A slight amount	A moderate amount	An intense amount
(0)	(1)	(2)	(3)

2. Have you felt any tremor (involuntary shaking of the hands, arms, or legs) during the last half hour or so?

Not at all	A slight amount	A moderate amount	An intense amount
(0)	(1)	(2)	(3)

B. For evaluation of the effects of chlorpromazine: Any direct measure (such as blood pressure) of the effects of the chlorpromazine injection on each subject was pretty much out of the question owing to limitations of time and personnel. It is known, however, that chlorpromazine does have somewhat of a dehydrating effect. As some indication that within the experimental time interval the chlorpromazine had been absorbed, the following questions were asked:

1. Does your mouth feel dry?

Not at all dry	A little dry	Somewhat dry	Very dry	Extremely dry
(0)	(1)	(2)	(3)	(4)

2. Does your nose feel stuffy?

Not at all stuffy	A little stuffy	Somewhat stuffy	Very stuffy	Extremely stuffy
(0)	(1)	(2)	(3)	(4)

Film Detail Questionnaire

Since it is known that chlorpromazine produces drowsiness, it seemed possible that experimental differences might be due to the fact that subjects had simply not watched the film. In order to check on this, a 10-item multiple-choice test concerned with small details of the film was administered. This test was rationalized to the subjects as a means of measuring the amount of time they had watched the movie, therefore, the amount of retinal stimulation received. Presumably the more they had watched the screen the more details they would remember.

Following this test the purpose of the experiment was disclosed, the deception was explained in detail, and the subjects were sworn to secrecy. Finally, the subjects filled out a brief questionnaire concerned with their past experiences with adrenalin and tranquilizers and with their suspicion, if any, of the experiment.

RESULTS

Physical Effects of the Injections

Evaluating, first, the effects of the injections, it can be seen in Table 1 that there are good indications that epinephrine has produced the required pattern of sympathetic activation. On self-reports of palpitation and tremor, subjects in the epinephrine condition report considerably more disturbance than subjects in either the placebo or chlorpromazine condition. On pulse measures, epinephrine subjects increase significantly when compared with placebo subjects. A subject's pulse was measured immediately before the injection and shortly after the movie. Pulse increased for some 63% of the epinephrine subjects and for 28% of the placebo subjects.

As to the effects of chlorpromazine, it can be seen in Table 1 that subjects in this condition report considerably more nose stuffiness and mouth dryness than subjects in the placebo or epinephrine conditions. This may be taken as indirect evidence that within the time limits of the experiment, chlorpromazine was taking effect. The increase in pulse rate (61% of chlorpromazine subjects increase) is a standard reaction to chlorpromazine and appears to be compensatory for the decreased blood pressure caused by this agent. It should be noted, however, that unlike epinephrine subjects the chlorpromazine subjects were unaware of this increased heart rate, for on the palpitation scale they are quite similar to subjects in the placebo condition.

Six subjects (included in Table 1) in the epinephrine condition were unaffected by the injection. They reported virtually no palpitation or tremor and their pulses were not markedly affected. Since for these subjects the necessary experimental state was not produced, they are not included in any further presentation of data.

TABLE 1

PHYSICAL EFFECTS OF THE INJECTIONS

Condition	N	Pulse		Palpitation	Tremor	Mouth dry	Nose stuffy
		Pre-injection	Post-injection				
Epinephrine	44	81.4	87.3	2.00	1.86	0.72	0.39
Placebo	42	78.7	75.5	0.30	0.12	0.30	0.68
Chlorpromazine	46	81.4	86.0	0.52	0.26	1.12	2.16
p value							
Epinephrine versus Placebo				<.001	<.001	<.01	*ns*
Epinephrine versus Chlorpromazine				<.001	<.001	.07	<.001
Placebo versus Chlorpromazine				*ns*	*ns*	<.001	<.001

Effects of Drowsiness

It is known that chlorpromazine produces drowsiness, a state which in this experimental context might mean that the subjects paid less attention to the movie. Differences between chlorpromazine subjects and those in the other two conditions in reaction to the film could then be due to differential attention rather than to the factors presumably being tested. In order to check for this, the film details questionnaire, described earlier, was administered shortly after the movie. The results of this questionnaire are presented in Table 2 where the figures represent the mean number of correct answers in each of the conditions. It can be immediately seen that the three conditions are virtually identical. None of these figures is significantly different from one another and any differences in reaction to the film cannot, then, be attributed to differences in attention.

Overt Reactions to the Film

The observation record provides a continuous record of each subject's reaction to the film. As an overall index of amusement, the number of units in which a subject's behavior was recorded in the categories Smile, Grin, Laugh, and Big Laugh are summed together. The means of this amusement index are presented in Table 3. The larger the figure, the more amusement was manifest. Differences are in the anticipated direction. Epinephrine subjects gave indications of greater amusement than did placebo subjects who, in turn, were more amused than chlorpromazine subjects. The *U* test was used to test for significance of differences since the variance in the epinephrine condition was significantly greater than that in either the placebo or chlorpromazine condition. The means of both the epinephrine and the placebo conditions are significantly greater than the mean of the chlorpromazine condition.

Though the trend is clearly in the predicted direction, epinephrine and placebo subjects do not differ significantly in this overall index. The difference between these two groups, however, becomes apparent when we examine strong reactions to the film. Considering just the categories Laugh and Big Laugh, as indicating strong reactions to the film, we find an average of 4.84 such units among the epinephrine subjects and of only 1.83 such units among placebo subjects. This difference is significant at better than the .05 level of significance. Epinephrine subjects tend to be openly amused at the film, placebo subjects to be quietly amused. Some 16% of epinephrine subjects reacted at some point with belly laughs while not a single placebo subject did so. It should be noted that this is much the state of affairs one would expect from the disguised injection of epinephrine

TABLE 2

MEAN NUMBER OF CORRECT ANSWERS ON THE FILM DETAILS QUESTIONNAIRE

Condition	N*	Mean number correct
Epinephrine	38	9.29
Placebo	41	9.15
Chlorpromazine	45	9.38

* One subject in the placebo and one in the chlorpromazine condition did not answer this questionnaire.

TABLE 3

THE EFFECTS OF EPINEPHRINE, PLACEBO, AND CHLORPROMAZINE ON AMUSEMENT

Condition	*N*	Mean amusement index
Epinephrine	38	17.79
Placebo	42	14.31
Chlorpromazine	46	10.41
p value		
Epinephrine versus Placebo		*ns*
Epinephrine versus Chlorpromazine		.01
Placebo versus Chlorpromazine		.05

—a manipulation which, as Schachter and Singer (1962) have suggested, creates a bodily state "in search of" an appropriate cognition. Certainly laughter can be considered a more appropriate accompaniment to the state of sympathetic arousal than can quietly smiling.

It would appear, then, that degree of overt amusement is directly related to the degree of manipulated sympathetic activation.

Evaluation of the Film

Responses to the postmovie questionnaire in which the subjects evaluated the film are presented in Table 4. The column heading Funny includes answers to the questions "How funny did you find this film?"; the heading Enjoy includes answers to the question "All in all, how much did you enjoy the film?"; and the heading Recommend presents answers to the question, "Would you recommend that we show this particular film to our future subjects?"

For all three questions, the trend is in precisely the same direction, epinephrine subjects like the film slightly more than do placebo subjects who like it more than do chlorpromazine subjects. On all questions, however, the differences between conditions are small and, at best, significant at only borderline probability levels.

TABLE 4

EVALUATION OF THE FILM

Condition	*N*	Funny	Enjoy	Recommend
Epinephrine	38	4.09	3.99	2.04
Placebo	42	4.01	3.95	1.93
Chlorpromazine	46	3.85	3.85	1.85

The fact that between-condition differences are large on the behavioral measure and quite small on the attitude scales administered after the film is an intriguing one. The most reasonable explanation comes from the subjects themselves. For example, after the experiment an epinephrine subject said,

> I just couldn't understand why I was laughing during the movie. Usually, I hate Jack Carson and this kind of nonsense and that's the way I checked the scales.

For this subject, then, his long time preferences determined his answers to the questions whereas his immediate bodily state seems to have determined his reaction while watching the movie. If this is widespread, it should be anticipated that there will be relatively little relationship between past preferences and overt behavior in the epinephrine condition and a considerably stronger relationship in the placebo condition. For chlorpromazine, too, one should anticipate slight relationship between past preferences and behavior. No matter what the long time feeling about such films the immediate reaction to the movie should be restrained owing to lack of sympathetic activity. However, as pointed out earlier, chlorpromazine is a weak blocker and, most reasonably, one should expect a somewhat weaker relationship with this drug than with the placebo.

TABLE 5

RELATIONSHIP BETWEEN LAUGHTER AND PREVIOUS ATTITUDE TOWARD SLAPSTICK FILMS

Attitude to slapstick	Epinephrine		Placebo		Chlorpromazine	
	N[a]	Laugh index	*N*[a]	Laugh index	*N*[a]	Laugh index
Dislike	25	15.12	29	10.52	29	7.34
Like	12	21.75	12	23.25	16	14.69
t		1.18		3.40		2.04
p		*ns*		.001		.05

[a] One subject in each condition did not answer the question concerned with general attitude to such films.

As a measure of general attitude toward the sort of film shown, at the time that they were evaluating the film the subjects also answered the question, "In general, how well

do you like this kind of slapstick film?" by checking one of five points along a scale ranging from "Slapstick is the kind of film I like least" to "Slapstick is my favorite kind of film." The relationship of attitude to this sort of film to reactions to this particular film in each of the drug conditions is presented in Table 5. The subjects are divided into two groups—those who dislike slapstick and those who like it as much or more than they do other kinds of films. The entries in the table are the mean laugh indices for each of the breakdowns.

It is evident that there is a very strong relationship between general attitude toward such films and laughter in the placebo condition, a considerably weaker relationship in the chlorpromazine condition, and the weakest relationship of all in the epinephrine condition.

DISCUSSION

The overall pattern of experimental results of this study and the Schachter and Singer (1962) experiment gives consistent support to a general formulation of emotion as a function of a state of physiological arousal and of an appropriate cognition. The fact that the epinephrine-placebo difference in this study, though in the proper direction, was not larger must be considered within the context of other relevant studies. As noted earlier, Schachter and Singer obtained similar results in their tests of the effects of epinephrine on euphoria and anger. In their attempt to account for their results, they identify two factors which could attenuate the differences between subjects injected with epinephrine and those receiving a placebo. One of these factors—the self-arousal of placebo subjects—has been tested in the present study. The second factor they identify is what they call the "self-informing" tendency of epinephrine subjects. To understand this notion will require a brief review of the formulation proposed by Schachter (1959) who has suggested that an emotion be considered a joint function of a state of physiological arousal and of a cognition appropriate to this state. Given a state of physiological arousal for which an individual has no immediate explanation, he will "label" this state and describe his feelings in terms of the cognitions available to him. Given a state of arousal for which an individual has a completely appropriate explanation (e.g., "I feel this way because I have just received an injection of adrenalin.") no evaluative needs will arise and he is unlikely to label his feelings in terms of the alternative cognitions available. These propositions are strongly supported in the Schachter-Singer study where subjects, injected with epinephrine and told precisely what they would feel and why, were considerably less emotional (either angry or euphoric) than were subjects injected with epinephrine and told simply that they would experience no side effects. Inevitably, however, some of the subjects in this latter condition were self-informed; that is, on their own, they attributed their states of arousal to the injection. Consistent with expectations, such "self-informed" subjects proved to be considerably less emotional than subjects in the same condition who were not self-informed. To the extent, however, that this self-informing tendency operates, the differences between placebo and epinephrine conditions will, then, be attenuated. There is little question that such a tendency also operated in the present study and we suggest, of course, that this is one of the chief factors limiting the magnitude of differences between the epinephrine and the placebo conditions. Such a self-informing tendency will probably operate in any experiment on humans which employs an injection technique.

In order to make the epinephrine-placebo comparison under conditions which would rule out the operation of any "self-informing" tendency, two experiments were conducted on rats. In one of these, Singer (1961) demonstrated that under fear inducing conditions, rats injected with epinephrine were considerably more frightened than rats injected with a placebo. In another study, Latané and Schachter (1962) demonstrated that rats injected with epinephrine were notably more capable of avoidance learning than were rats injected with a placebo. Viewed together, this series of experiments on rats and humans give clear support to the hypothesis that "emotionality" is, in part, a function of degree of sympathetic excitation.

The identification of the "self-informing" tendency does permit us to consider one alternative interpretation of the results of the experiment presented in this paper. One might consider that the effects of the several injections have been to vary "level of activation" in the sense employed by Lindsley (1951) and Woodworth and Schlosberg (1958). This would imply that no matter what the state or activity, epinephrine subjects would react more extremely than either placebo or chlorpromazine subjects. If correct, this would negate the cognitive component of this formulation of emotion. It must be remembered, however, that this experiment was planned hand in hand with the Schachter and Singer (1962) study. In this prior study the "epinephrine informed" conditions were deliberately built into the experiment to test for this possibility. Subjects who were injected with epinephrine and told precisely what they would feel and why did not "catch" the induced emotional state at all. It seems safe to generalize from these results to the present study and conclude that the level of activation notion alone cannot explain the results of these two studies.

SUMMARY

An experiment is described which was designed to test the proposition that "emotionality" is, in part, a function of the degree of excitation of the sympathetic nervous system. The degree of sympathetic activation was manipulated by injections of (*a*) the sympathomimetic agent—epinephrine, (*b*) a placebo, and (*c*) the sympatholytic drug—chlorpromazine. The effects of these drugs on amusement were tested by exposing subjects to a slapstick film. Epinephrine subjects were more amused than were placebo subjects who, in turn, were more amused than chlorpromazine subjects.

REFERENCES

KILLAM, EVA K. The pharmacological apsects of certain drugs useful in psychiatry. In, *Psychopharmacology: Problems in evaluation*. Washington, D. C.: National Academy of Sciences, National Research Council, 1959. (NAS-NRC Publ. No. 583)

LATANÉ, B., & SCHACHTER, S. Adrenalin and avoidance learning. *J. comp. physiol. Psychol.*, 1962, **55**, 369–372.

LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 473–516.

SCHACHTER, S. *The psychology of affiliation*. Stanford, Calif.: Stanford Univer. Press, 1959.

SCHACHTER, S., & SINGER, J. E. Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev.*, 1962, in press.

SINGER, J. E. The effects of epinephrine, chlorpromazine and dibenzyline upon the fright responses of rats under stress and non-stress conditions. Unpublished doctoral dissertation, University of Minnesota, 1961.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1958.

(Received June 23, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 129–131

ORDINAL POSITION AND CONFORMITY

SELWYN W. BECKER AND JEAN CARROLL

Graduate School of Business, University of Chicago

Schachter (1959) has posited a need for affiliation and presented corroborative data in support of his contention. He also has cited evidence that this need is stronger in first born than in later born children. Since an affiliative need is satisfied in the presence of others, an individual wishing to satisfy this need might seek to join a group. On becoming a group member, possibly one exposes himself to normative social influence. Such influence attempts are usually directed toward a nonconforming member and are designed to align him more closely with the group norm. Prolonged resistance can lead to eventual rejection and isolation (Schachter, 1951). Presumably, once he is a group member an individual high in need Affiliation would avoid actions tending to isolate him from the group, his source of satisfaction.

Striving for group membership, or adoption of a reference group, also can be viewed as a desire to affiliate with that group. Such striving can result in the adoption of that group's norms (Siegel & Siegel, 1957) or by inference, a willingness to accept that group's definition of social reality. Again, presumably, in the presence of members of his reference group, an individual will avoid actions which might preclude his eventual acceptance as a member.

The Asch (1956) situation wherein subjects exhibit varying degrees of conformity to a majority judgment also has been viewed as a situation where such yielding is a manifestation of susceptibility to normative social influence (Deutsch & Gerard, 1955).

Based on these considerations this experiment investigated (*a*) the greater susceptibility of persons high on need Affiliation to normative social influence compared to those low on need Affiliation and (*b*) the greater susceptibility of those aspiring to group membership to normative social influence than established members.

METHOD

Forty-eight boys who play regularly at the same public playground served as subjects, for an exact replication of the Asch (1956) study where a minority of one (naive subject) was exposed to the unanimous judgment of three native American accomplices. Each subject made 18 judgments, 12 of which were "critical" trials designed to measure yielding behavior.

Need Affiliation. Of the 48 subjects, 24 were the first born children in their families and 24 later born. The first borns were considered the high need Affiliation group, and the later borns the low need Affiliation group.

Reference Group. Eighteen of the 48 subjects were native born Puerto Ricans and all spoke fluent English. For purposes of the experiment it was assumed that the Puerto Rican children had adopted the native American children as their reference group.[1] Ten of the Puerto Rican group were the first born children in their families and eight were later borns.

With these samples it was possible operationally to test the hypotheses:

Hypothesis I. In the Asch situation first born children will yield more frequently than will later born children.

Hypothesis II. In the Asch situation children of Puerto Rican extraction will yield more frequently than will native American children.

Because the Puerto Rican group also was divided according to birth order, it was possible to test a subsidiary hypothesis based on the notion that birth order and reference group effects together should be stronger than either effect by itself.

Hypothesis III. In the Asch situation first born Puerto Ricans will yield more frequently than will later born Puerto Ricans or first born native Americans, who, in turn, will yield more frequently than will later born native American children.

RESULTS AND DISCUSSION

The data[2] appropriate for a test of the first hypothesis are summarized in Table 1.

[1] In connection with a succeeding study, the 18 Puerto Ricans ranked their classmates in order of friendship potential. Eight named native American boys as their first choice, 10 named one of two Puerto Ricans tied for the highest score on measures of acculturation as their first choice.

[2] Detailed data (Tables A, B, C, and D), on which the tables in this report are based, have been deposited with the American Documentation Institute. Order Document No. 7193 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25

TABLE 1

YIELDING BEHAVIOR IN HIGH AND LOW NEED AFFILIATES

Birth order	N	Number of errors	Percentage yielding	p
First born	24	117	40.6	<.01
Later born	24	85	29.5	

Differences in yielding behavior between first borns and later borns were found to be significantly different from what might be expected by chance ($p < .01$) by the use of the Mann-Whitney U test (Siegel, 1956). It is concluded that those high on need Affiliation more frequently yield to normative social influence than those low on need Affiliation.

The second hypothesis, that individuals striving for membership in a group will be more susceptible to normative social influence than will be members of that group, is also supported by the data, summarized in Table 2. These data were also tested for significance with the Mann-Whitney U test (Siegel, 1956). The observed differences in yielding errors were in the predicted direction and significantly different from chance ($p < .01$).

The third hypothesis, concerning the combined reference group and need Affiliation effects, was stated in an ordered form, i.e., that one group would be higher than the second, which would be higher than the third. In the absence of any data on the relative strengths of need Affiliation and reference group effects, no distinction was made between the later born Puerto Rican and the first born American groups so these two groups were combined for the test of Hypothesis III. The ordering of the three groups was correctly predicted and the difference in yielding errors was found to be significant ($p < .01$) by the use of the Jonckheere (1954) test, so it is concluded that the hypothesis is supported by the data.

However, as stated, the hypothesis explicitly assumes that need Affiliation is differential in both the Puerto Rican and American groups, thus, enabling one to predict the ordering of three of the four groups. If, however, the reference group effect was the only one operating in the Puerto Rican group, and furthermore if it induces greater yielding than does high need Affiliation, then it is possible that the conclusion supporting Hypothesis III is artifactual due to the different group sizes and the strength of the reference group effect.

In order to resolve this question, differences in yielding behavior between first borns and later borns separately was examined for the two ethnic groups. These data are presented in Table 4.

Based on the Mann-Whitney U test, it cannot be concluded that birth order makes a significant difference in the Puerto Rican group. Thus, doubt is cast on the conclusion drawn with regard to the relation between reference groups, need Affiliation, and their combined effects on yielding to normative

for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 2

YIELDING BEHAVIOR IN THE PRESENCE OF REFERENCE GROUP MEMBERS

Group	N	Number of errors	Percentage yielding	p
Puerto Rican	18	104	50.9	<.01
Native American	30	98	27.2	

TABLE 3

RELATIONS BETWEEN NEED AFFILIATION, REFERENCE GROUP, AND YIELDING BEHAVIOR

Group and birth order	N	Number of errors	Percentage yielding	p
Puerto Rican first born	10	60	50.0	<.01
Puerto Rican later born and American first born	22	101	38.25	
American later born	16	41	21.35	

TABLE 4

BIRTH ORDER AND YIELDING IN PUERTO RICAN AND AMERICAN GROUPS

Birth order	N	Number of errors	Percentage yielding	p
Puerto Rican				
First born	10	60	50.0	>.05
Later born	8	44	45.8	
American				
First born	14	57	33.9	<.01
Later born	16	41	21.35	

social influence. Such doubt, however, need only be associated with the present test of the hypothesis.

One can also consider the reasons why birth order and affiliative need might be unrelated in cultures other than our own. If, as Schachter (1959) assumes, the relation between birth order and need Affiliation is due to conditioning procedures specifically related to being the first or only born child in a family group, and since child rearing practices are not common to all cultures, it is not surprising to find such a relationship lacking in children of another culture. These conjectures, however, must remain unresolved due to the small sample sizes of 8 and 10 Puerto Rican children.

The findings that birth order and the presence of reference group members partially determine acceptance of social influence have implications for the formation and structure of groups. One might expect, for instance, that first borns join a greater number of social groups than do later borns, or that in any given group more first borns will be members than might be expected by chance.

Other expectations might be associated with the leadership function in groups. If leaders reflect the mean attitude of the group members, then perhaps a preponderance of leaders will be first born or only children. On the other hand, if leaders arise as a function of their initiating or less conforming behavior, fewer first borns might be leaders.

It was felt that some ad hoc data relevant to these sorts of speculations might be of some interest. To that end, data gathered for a research project [3] to be reported elsewhere, was examined. The project is concerned with delinquency and juvenile gangs. The most powerful leader in five large, delinquent, Chicago area gangs was identified and four of the five were first born children in their families. The five gangs have a total membership of 211 of which 97, or 46%, are the first born or only children in their families. While insufficient data are available for conclusions with any certainty, it does not seem unlikely, especially since these groups are from the lowest socioeconomic classes, that the percentage of first born gang members exceeds the percentage of first borns in the population. Whether this relationship holds for different kinds of groups, or in groups from different socioeconomic classes, is of course, subject to question. Need Affiliation as a determiner of group formation and structure appears to offer an interesting and fruitful avenue for further investigation.

[3] The project, Youth Studies Project, is supported by the United States Public Health Service and directed by J. F. Short, Jr., Department of Sociology, the University of Chicago.

SUMMARY

Forty-eight boys served as subjects for an exact replication of the Asch (1956) situation. Twenty-four were the first born children in their families and 24 later born. First borns were considered the high need Affiliation group, and later borns the low need Affiliation group. Eighteen members of the group were native born Puerto Ricans who, it was assumed, had adopted the native American children as their reference group.

The hypotheses that high need Affiliation and aspiration to group membership would be associated with greater conformity in the Asch (1956) situation were supported by the data.

REFERENCES

ASCH, S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**(9, Whole No. 416).

DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.

JONCKHEERE, A. R. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.

SCHACHTER, S. *The psychology of affiliation.* Stanford: Stanford Univer. Press, 1959.

SIEGEL, A., & SIEGEL, S. Reference groups, membership groups, and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 360–364.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received June 27, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 132–135

CRITIQUE AND NOTES

A DEVELOPMENTAL STUDY OF DOGMATISM DURING ADOLESCENCE WITH REFERENCE TO SEX DIFFERENCES[1]

C. C. ANDERSON

University of Alberta

Loevinger (1958) has described a major aspect of ego development termed "the ability to assume distance from oneself, or . . . capacity to conceptualise oneself" (p. 36), which consists of two components, nonegocentricity or "distance from the ego" and nonimpulsivity or "distance from impulses." She postulates a continuum running from childhood, where both components are absent, to adulthood, where both are present and a relatively realistic self-concept has developed which can be measured by self-report devices. The adolescent is at the midpoint of this continuum, having achieved "distance from impulses"; without an equal development of the other component, however, his self-appraisal can be of only a stereotyped sort so characteristic of dogmatism. The primary hypothesis of the present research is that there is a progressive decline in dogmatism with advancing age from Grades 8 to 12.

The second hypothesis tested was concerned with sex differences in dogmatism. Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) report no consistent sex differences in performance on scales designed to measure aspects of authoritarianism (pp. 138, 173–174, 178) and this finding has been replicated by Stotsky (1955). However, Bendig and Hountras (1959) and Irvine (1957) find that undergraduate males in psychology score significantly higher on the F test than undergraduate females. Both findings—that there are no sex differences and that males are significantly more dogmatic than females—run counter to currently accepted theory (McCandless, 1961) concerning the origins of dogmatism. Block (1955) has demonstrated that a major determinant of dogmatism is a pattern of restrictive child rearing practices which, according to Sears, Maccoby, and Levin (1957), are more often directed at the young girl than the young boy, and which, according to Jackson (1956), are carried out less often by the father than by the mother with whom the girl is most likely to identify (Kagan, 1956; Maccoby & Wilson, 1957). The vicious circle promoting female dogmatism is likely to be strengthened by the fact that females are generally dissatisfied with their standard submissive secondary role in contemporary masculine oriented cultures (Brown, 1958; Lynn, 1959). On the basis of this theory, therefore, it would be predicted that females are significantly more dogmatic than males, a prediction given some support by Kates and Diab (1955), who found that the correlation between child rearing attitudes and dogmatism for female subjects was significant while that for males was not significant.

A possible solution to this problem involves the assumption that intelligent females are significantly more hostile than duller females both to the restrictive child rearing practices and to the privileges accorded to the males in contemporary society. If this assumption is correct, then it becomes possible to state a second and third hypothesis: second, that there are no sex differences in dogmatism, and third, that intelligent females are significantly more dogmatic than equally intelligent males, and relatively duller males are significantly more dogmatic than relatively duller females.

METHOD

A representative sample of 290 Grade 8 students in the prairie city of Edmonton, Alberta, took a version of Rokeach's (1960) Dogmatism Scale abbreviated to 26 items which were judged by the present writer to be suited to their level of maturity. The test, administered orally by the writer and his graduate students and a few interested teachers, was also given to a group of students in a composite high school. Four classes were tested in Grade 10, five in Grade 11, and six in Grade 12, the classes being selected at each grade level so as to yield a roughly normal spread of ability. Intelligence test scores were made available,[2] CTMM Short Form Verbal IQs for the Grade 8 students and Otis AM IQs for the high school students. Other scores at hand for the Grade 8 students included an estimate of socioeconomic status based on Gough (1949) and

[1] This investigation was supported by a research grant from the University of Alberta.

[2] By W. B. Elley, of the University of Alberta, and A. Kingsley Dean, of Bonnie Doon Composite High School, Edmonton.

TABLE 1

AVERAGE AND SPREAD OF DOGMATISM SCORES AT FOUR ACADEMIC LEVELS

Scores	Grades			
	8	10	11	12
Average	40.80	37.28	35.59	32.38
SD	15.18	13.10	14.15	14.00
N	280	142	163	183

a measure of the capacity to control impulse gratification which had been used by Luft (1957) and Shippee-Blum (1959).

RESULTS

Negative scores on the Dogmatism Scale were eradicated by adding a constant +40 to all scores, and the first hypothesis was tested by determining the significance of the difference between average dogmatism scores at all grade levels tested, the basic data for this appearing in Table 1. The following mean differences are significant: Grades 8–11, 8–12, 10–12, and 11–12. Clearly Loevinger's (1958) hypothesis is verified.

The second hypothesis was tested by determining the significance of the difference between the average dogmatism scores of females and males at all academic levels, the basic data for this appearing in Table 2.

At three of the four grade levels (8, 11, and 12) the females are consistently but not significantly more dogmatic than males; at the remaining grade level (10), the females are slightly but not significantly less dogmatic. A decision about the second hypothesis, that there are no sex differences in dogmatism, cannot be made with certainty on the basis of these results.

The third hypothesis was validated by dividing females and males independently, at each academic level, into an upper and a lower group in terms of intelligence test scores, approximately 20% of the grade sample appearing in each group, and by determining the significance of the Sex × Intelligence interaction term in an analysis of variance of dogmatism scores categorized dichotomously by sex and intelligence. *F* ratios for all relevant sources of variance are presented in Table 3. There is consistent and, at Grades 11 and 12, significant evidence of an interaction between sex and intelligence, the nature of which is revealed in Table 4. The evidence presented in the last two tables (Tables 3 and 4) is consistent with the view that intelligent females tend to be significantly more dogmatic than intelligent males, while the accuracy of the converse position, that relatively dull males are significantly more dogmatic than relatively dull females, cannot be decided upon from the present data. Indeed, the state of the decision to be made with respect to the second part of the third hypothesis is identical with that to be made to the second hypothesis, with which it is logically related, on the basis of the data in Table 2.

It should be noted that the increasing difference with age between the average dogmatism scores of intelligent and of relatively dull groups is probably due to the progressive decline with age of the average dogmatism score of the intelligent groups coupled with the steady maintenance of a high average dogmatism score by the relatively dull groups.

A further prediction can be made to test the

TABLE 2

AVERAGE AND SPREAD OF DOGMATISM SCORES FOR MALES AND FEMALES AT FOUR ACADEMIC LEVELS

Scores	Grades							
	8		10		11		12	
	F	*M*	*F*	*M*	*F*	*M*	*F*	*M*
Average	41.38	40.34	36.68	38.03	37.47	33.92	33.00	32.00
SD	16.25	14.04	12.80	13.45	14.67	13.50	13.37	14.41
N	125	155	79	63	75	88	80	103

TABLE 3

F RATIO DERIVED FROM AN ANALYSIS OF VARIANCE OF DOGMATISM SCORES CATEGORIZED BY SEX AND INTELLIGENCE

Source	Grades			
	8	10	11	*F*
Sex	.038	.179	2.167	.392
Intelligence	1.190	.964	3.144	3.716
Sex × Intelligence	5.830	1.143	6.253*	5.243*

* Significant beyond the .05 level.

TABLE 4

AVERAGE DOGMATISM SCORES OF UPPER AND LOWER INTELLIGENCE GROUPS OF FEMALES AND MALES AT FOUR ACADEMIC LEVELS

Group	Grades							
	8		10		11		12	
	F	*M*	*F*	*M*	*F*	*M*	*F*	*M*
Upper	41.43	37.86	34.87	33.33	37.07	27.33	35.27	29.87
Lower	39.62	43.21	36.93	38.53	40.27	38.27	37.27	38.93

TABLE 5

ANALYSIS OF VARIANCE OF DOGMATISM SCORES CATEGORIZED BY SEX, INTELLIGENCE, AND ANXIETY LEVELS

Source	*df*	*MS*	*F*
Sex	1	227.82	1.144
Intelligence	1	148.52	.746
PCQ	1	787.52	3.953
Sex × Intelligence	1	503.17	2.526
Sex × PCQ	1	431.57	2.166
Intelligence × PCQ	1	113.87	.572
Sex × Intelligence × PCQ	1	1,486.48	7.461*
Residual	72	14,344.44	
Total	79		

* Significant beyond the .05 level.

validity of the present interaction hypothesis. Rokeach (1960, pp. 348–349) has demonstrated that one correlate of dogmatism is anxiety which he traces to its origin in the sort of restrictive upbringing already mentioned (p. 359). If this is correct and if the second hypothesis, particularly the part which states that intelligent females are significantly more dogmatic than intelligent males, is also correct, then a necessary prediction will take the following form: that an analysis of variance of dogmatism scores categorized by sex, intelligence, and anxiety levels, will yield at least a significant second-order interaction of Sex × Intelligence × Anxiety, the intelligent anxious females being significantly more dogmatic than the intelligent nonanxious females, and the intelligent anxious males being significantly more dogmatic than the intelligent nonanxious males, with a corresponding though not so significant trend appearing in the case of the duller groups. Rokeach's measure of anxiety was one developed by Welch (1952). The present writer had available, for the Grade 8 sample, scores on the PCQ, a similar sort of measure scaled in the opposite direction, a high PCQ score signifying a small amount of anxiety. Ten students were chosen to represent each of the eight categories (two sexes, and two intelligence and two anxiety levels) and an analysis of variance of the dogmatism scores categorized in this way is presented in Table 5. The significant interaction term indicates that anxiety is associated with the formerly established Sex × Intelligence factor in dogmatism. The average dogmatism scores for all eight categories made this clear. These were: intelligent females high in anxiety 57.4, intelligent females low in anxiety 35.4, intelligent males high in anxiety 37.5, intelligent males low in anxiety 31.1, dull females high in anxiety 41.0, dull females low in anxiety 39.8, dull males high in anxiety 43.5, and dull males low in anxiety 42.0. Therefore the present findings with respect to Sex × Intelligence differences in dogmatism are congruent with those of Rokeach with respect to anxiety and dogmatism.

A final point can be established which Rokeach (1960) did not take account of. An analysis of the scores of the intelligent male and female groups shows a significant negative relationship between socioeconomic status and anxiety, and between socioeconomic status and dogmatism. The first relationship can be demonstrated by an analysis of variance of the socioeconomic indices of 40 intelligent males and females categorized by sex and anxiety level. This yielded a significant Sex × Anxiety interaction with an F of 18.57, significant at the 1% level, $df = 1/36$. Specifically, the average measure of socioeconomic status for 10 intelligent females low in anxiety (15.9) was significantly greater ($t = 2.102$) than that for 10 intelligent females high in anxiety (13.00), and the average measure of socioeconomic status for 10 intelligent males low in anxiety (17.1) was significantly greater ($t = 2.567$) than that for 10 intelligent males high in anxiety (14.1). The second relationship between socioeconomic status and dogmatism can be demonstrated by an analysis of variance of the socioeconomic indices of 40 intelligent males and females categorized by sex and dogmatism level. This yielded a significant Sex × Dogmatism interaction with an F ratio of 22.11, significant at the 1% level, $df = 1/36$. Specifically, the average measure of socioeconomic status for 10 intelligent nondogmatic females (17.00) was significantly greater ($t = 3.362$) than that for 10 intelligent dogmatic females (13.8) and the average measure of socioeconomic status for 10 intelligent nondogmatic males (16.8) was greater than that for 10 intelligent dogmatic males (14.00), the difference in this case being insignificant ($t = 1.643$). Unfortunately the interaction between socioeconomic status and the other factors in dogmatism could not be estimated because the present data did not yield enough dull children who came from the upper socioeconomic group.

DISCUSSION

These results make trouble for Davis' (1948) conception of parentally instigated "socialized anxiety" as the prime source of motivation of middle-class children. The child rearing practices reported to underlie such anxiety are similar to those Frenkel-Brunswik (1954, pp. 236–237) describes as characteristic of authoritarian parents, and yet intelligent children of superior socio-

economic status are significantly less dogmatic and less anxious than intelligent children of inferior socioeconomic status. It may well be that this finding can be generalized to the complete sample since very few children of superior status are in the lower intelligence group. Clearly the lower-class children in this sample are a great deal more anxious than Davis would have us believe.

The results also lend some plausibility to Adorno's (Adorno et al., 1950) hypothesis that dogmatism is associated with upward social mobility because the present intelligent dogmatic males coming from families of low socioeconomic status show the same characteristics as those in Douvan and Adelson's (1958) second pattern of upward mobility—intelligence, dogmatism, and anxiety.

SUMMARY

An abbreviated version of Rokeach's (1960) Dogmatism Scale was administered to samples of children in Grades 8, 10, 11, and 12. Intelligence test scores were also available for all students, and measures of socioeconomic status and psychological control (the inverse of anxiety) for the Grade 8 students only. The first hypothesis, that there is a significant decline in dogmatism during the adolescent years, was verified. No decision could be arrived at about the validity of the second hypothesis, that there are no sex differences in dogmatism. The third hypothesis, that Sex × Intelligence is a significant interaction component of the variance of dogmatism scores, was verified, the tendency for intelligent females to be significantly more dogmatic than intelligent males being especially prominent. Evidence was also provided establishing anxiety and socioeconomic status as two more correlates of dogmatism. The former relationship between anxiety and dogmatism linked the third hypothesis with Rokeach's work, and the latter inverse relationship linking socioeconomic status with dogmatism makes plausible the inference that child rearing practices are basic determinants of dogmatism.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

BENDIG, A. W., & HOUNTRAS, P. T. Anxiety, authoritarianism, and student attitudes towards departmental control of college instruction. *J. educ. Psychol.*, 1959, **50**, 1–8.

BLOCK, J. Personality characteristics associated with fathers' attitudes towards child-rearing. *Child Develpm.*, 1955, **26**, 41–48.

BROWN, D. G. Sex-role development in a changing culture. *Psychol. Bull.*, 1958, **55**, 232–242.

DAVIS, A. *Social-class influences upon learning*. Cambridge: Harvard Univer. Press, 1948.

DOUVAN, E., & ADELSON, J. The psychodynamics of social mobility in adolescent boys. *J. abnorm. soc. Psychol.*, 1958, **56**, 31–44.

FRENKEL-BRUNSWIK, ELSE. Further explorations by a contributor to *The Authoritarian Personality*. In R. Christie & Marie Jahoda (Eds.), *Studies in the scope and method of The authoritarian personality*. Glencoe, Ill.: Free Press, 1954. Pp. 226–275.

GOUGH, H. G. A short social status inventory. *J. educ. Psychol.*, 1949, **40**, 52–56.

IRVINE, L. F. Sex differences and the relationship between certain personality variables. *Psychol. Rep.*, 1957, **3**, 595–597.

JACKSON, P. W. Verbal solutions to parent-child problems. *Child Develpm.*, 1956, **27**, 339–349.

KAGAN, J. The child's perception of the parent. *J. abnorm. soc. Psychol.*, 1956, **53**, 257–258.

KATES, S. L., & DIAB, L. N. Authoritarian ideology and attitudes on parent-child relationships. *J. abnorm. soc. Psychol.*, 1955, **51**, 13–16.

LOEVINGER, J. A theory of test response. In, *Invitational conference on testing problems*. Princeton, N. J.: Educational Testing Service, 1958.

LUFT, J. Psychological control patterns within families. *J. consult. Psychol.*, 1957, **21**, 206.

LYNN, D. B. Sex differences in masculine and feminine identification. *Psychol. Rev.*, 1959, **66**, 126–135.

MCCANDLESS, B. *Children and adolescents: Behavior and development*. New York: Holt, Rinehart, & Winston, 1961.

MACCOBY, ELEANOR E., & WILSON, W. C. Identification and observational learning from films. *J. abnorm. soc. Psychol.*, 1957, **55**, 76–88.

ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.

SHIPPEE-BLUM, EVA M. The young rebel: Self-regard and ego ideal. *J. consult. Psychol.*, 1959, **23**, 44–50.

STOTSKY, B. A. The authoritarian personality as a stereotype. *J. Psychol.*, 1955, **39**, 325–328.

WELCH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65–72.

(Received May 31, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 136–141

PROPHECY FAILS AGAIN:

A REPORT OF A FAILURE TO REPLICATE [1]

JANE ALLYN HARDYCK [2] AND MARCIA BRADEN

Stanford University

On July 4 of a recent year, a group of 135 men, women, and children vanished from their homes in a small southwestern town. Their homes were sealed; the windows were covered with newspapers; the cluster of houses was deserted. The only message they had left was a sign on the door of their church, reading "Gone for two weeks, camp meeting."

The neighbors of the group and the town officials soon discovered where the members of the Church of the True Word [3] had gone. In response to prophecies of a forthcoming nuclear disaster, the group had for many months been building and stocking underground fallout shelters, with as much secrecy as possible. On July 4, one of their prophets received a message, "The Egyptians are coming; get ye to the safe places," and they immediately obeyed what they believed to be a command from God. They were huddled in their shelters, awaiting the nuclear catastrophe. For 42 days and nights they remained there, in expectation of imminent disaster. While they stubbornly sat underground, the authors walked around the hot, dusty desert town piecing together the history of the group from interviews with townspeople and the few group members who, disillusioned, left the shelters.

The Church of the True Word is an evangelical Christian church associated with the Pentecostal movement. Its members believe in the Bible as the literal word of God and accept as operating today the gifts of the Holy Spirit delineated in First Corinthians of the New Testament, Chapters 12 and 14. These gifts include speaking and interpreting tongues, personal prophecy, and healing by faith. The titular head of the group, a Mrs. Shepard, is their minister and chief prophet, although important decisions are made only after she has consulted with two of the group members, Peter Jameson and David Blake, both of whom are also ordained ministers.

The "colony," as they call themselves, springs from two main sources. Mrs. Shepard established a following about 5 years ago in the small southwestern town, and soon began work on the present church building. In this, she had the help of Jameson and Blake, who at the time were missionaries to Central America from two congregations in the Middle West. The second source of members of the Church of the True Word was these midwestern churches.

Even in the early days of her ministry, Mrs. Shepard was preaching preparedness for nuclear attack, and almost 4 years ago a prophecy was received in the Midwest to the effect that "in fewer years than I have fingers on my right hand" there would be nuclear devastation. The more recent history of the Church of the True Word began with another prophecy. On November 23, 8 months before the group finally went underground, a prophet in the Midwest received word that "you have 6 months to prepare." On receiving this message, Blake, Jameson, and various others packed up and moved to the Southwest and about February began to build fallout shelters and homes.

The shelters were built "through the inspiration of God," according to the specifications of Civil Defense, which is, for these people, "the Noah's Ark of today." They were not designed as bomb shelters, since the group believed that their town would receive only fallout from a direct hit on Desert City, which lay 50 miles to the west. There were probably five large shelters under houses and four smaller ones dug in an open field nearby. The larger ones were quite livable, although far from luxurious, as they

[1] This investigation was carried out during the tenure of a Predoctoral Research Fellowship from the National Science Foundation to the first author and a Predoctoral Research Fellowship from the National Institute of Mental Health, United States Public Health Service, to the second author. The collection of data was undertaken in cooperation with the Studies in International Conflict and Integration, Stanford University, directed by Robert North. The authors would like to thank Larry Robertson, who helped in collecting the data, and Leon Festinger, who helped by his criticisms of many drafts of this paper.

[2] Now at the University of California, Survey Research Center.

[3] This is a fictitious name which, we believe, captures the flavor of the actual name of the group. All other names and places used in this paper have been similarly disguised.

lacked modern plumbing and were rather badly ventilated. The shelters were stocked with canned and dehydrated food, large cans of water, and other necessities, and were provided with generators for use when public power failed.

For the group to make such careful and extensive preparations, they must have had a rather clear and specific idea of what was to happen. Indeed they did. From a particular interpretation of portions of the book of Revelation, they believed that about one-third of the population of the earth would be wiped out by nuclear warfare and that injuries and sickness would be widespread among the survivors. The members of the Church of the True Word also expected that after the disaster they would receive special powers from God so that they might perform miracles of healing beyond what they were already able to do, and might be enabled to spread the gospel to all nations within the short space of about a year. It was the necessity of saving themselves for this purpose that dictated that they must keep their preparations secret. They feared that if the location and nature of their shelters were generally known, they would be unable to prevent others from breaking in at the time of the attack, thus creating a situation in which no one could survive.

From February until the "deadline" of May 23, many more families from the midwest congregations arrived to join in the preparations. There were also, of course, several families from the local community who were members of the church and who helped in the work. The shelters were not finished by May 23, and much apprehension arose among the members of the group. In an anxious flurry of preparations they waited until July 4, when they received the message, we believe through Mrs. Shepard, "The Egyptians are coming; get ye to the safe places." They then entered the shelters—29 families, about 135 men, women, and children.

Hypothesis to be Tested

Our interest in the True Word group arose because of their apparent similarity to the "doomsday groups" discussed by Festinger, Riecken, and Schachter (1956) in *When Prophecy Fails*. The historical accounts of such groups as well as an empirical study of a more recent group, the Lake City group, suggest that the failure of the members to confirm their pessimistic predictions led them to increase in fervor of belief and in proselyting. Festinger et al. state five conditions that they feel must be met for this to occur:

> 1. A belief must be held with deep conviction and it must have some relevance to action, that is, to what the believer does or how he behaves.
>
> 2. The person holding the belief must have committed himself to it; that is, for the sake of his belief, he must have taken some important action that is difficult to undo. In general, the more important such actions are, and the more difficult they are to undo, the greater is the individual's commitment to the belief.
>
> 3. The belief must be sufficiently specific and sufficiently concerned with the real world so that events may unequivocally refute the belief.
>
> 4. Such undeniable disconfirmatory evidence must occur and must be recognized by the individual holding the belief.
>
> The first two of these conditions specify the circumstances that will make the belief resistant to change. The third and fourth conditions together, on the other hand, point to factors that would exert powerful pressure on a believer to discard his belief.
>
> 5. The individual believer must have social support. It is unlikely that one isolated believer could withstand the kind of disconfirming evidence we have specified. If, however, the believer is a member of a group of convinced persons who can support one another, we would expect the belief to be maintained and the believers to attempt to proselyte or to persuade nonmembers that the belief is correct (p. 4).[4]

These five conditions define a situation in which the believer has two sets of cognitions that clearly do not fit together. That is, he is experiencing a great deal of dissonance between the cognitions corresponding to his belief and the cognitions concerning the failure of the predicted event to occur. This situation, however, is one in which it is almost impossible for the individual to reduce his dissonance. He cannot give up his strongly held beliefs, and he cannot deny that the predicted event has failed to occur. He is also unable either to reduce the importance of his commitment to his beliefs or to make the disconfirmation irrelevant to them. Therefore, the believer who holds to his belief under these conditions has but one recourse if he is to reduce the dissonance; he must seek new information consonant with his beliefs. One of the best sources of new consonant cognitions is the knowledge that others' beliefs are the same. The authors suggest, then, that the need for new supporting cognitions will lead the believer to try to convince others of the validity of his beliefs.

[4] From: *When Prophecy Fails* by Leon Festinger, Henry W. Riecken, and Stanley Schachter. University of Minnesota Press, Minneapolis. Copyright 1956 by the University of Minnesota.

Suitability of the True Word Group for a Test of the Hypothesis

Our purpose in learning about the history and beliefs of the True Word group was to determine whether the group met the conditions enumerated in *When Prophecy Fails* and thus would provide a test of the hypothesis under consideration. The first condition is that the group members must hold their belief with deep conviction and that the belief have some relevance to action. It is quite clear that the members held their general religious belief system with deep conviction. Many were originally ministers or missionaries actively engaged in Christian work, and most of the members to whom we spoke would refer to "gifts" they themselves or members of their families possessed. Also, as far as we were able to discover, Mrs. Shepard was respected by all of the congregation as a truly exceptional prophet. Thus, since the prophecy probably came from her and was closely tied to their belief system, it seems clear that it would be very strongly believed by the majority of the congregation. The obvious fact that the group had acted on their belief by building and entering the shelters is the strongest evidence for their belief in the prophecy and also, of course, proof that the belief had relevance for action.

The second condition is that the person holding the belief must have committed himself to it by some action difficult to undo. For the Midwest contingent, the commitment was extreme. They had given up their jobs, had picked up and moved over a thousand miles, and had invested a great deal of time, effort, and money in the building and stocking of homes and shelters. Those from the local area had perhaps given up less, but in several cases they also had lost jobs and had invested considerable sums of money. The things that they have done they cannot undo; the money is spent and the jobs are lost. Most important, none of them can deny or take back the fact that he spent 42 days in hot, humid, crowded shelters and he did this because of his belief.

The third condition, that the belief must be sufficiently specific and sufficiently concerned with the real world so that events may unequivocally refute it, is also quite easy to document. At the time that the group went into the shelters, they believed that a nuclear attack was imminent, and that they would not come out of the shelters until that attack had occurred. That is, they expected to return to a world that had been devastated.

The fourth point, that "undeniable disconfirmatory evidence must occur and must be recognized by the individual holding the belief," is also clearly met. No nuclear attack occurred while the group was in the shelters, and they did not return to a devastated world. Thus, we must conclude that the True Word group suffered the unequivocal disconfirmation of a specific prediction.

The last condition that must be met in order that the True Word group may provide an adequate test of the hypothesis is that the individual believer have social support. This was so clearly the case that it hardly needs documentation. The members of the group had been living together as a separate, rather isolated community for several months prior to July 4. Indeed, some of the members had known each other for years and many were related by blood or marriage. During the time of the disconfirmation, social support was not only present, it was unavoidable. There were as many as 35 people in each shelter, and the shelters provided absolutely no privacy. Furthermore, the shelters were organized in such a way that each contained at least one very strongly convinced member who could hold his group together, and all of the shelters were connected by an intercom system so that the leaders could be consulted in case any members should begin to weaken.

It can be concluded, then, that the five conditions enumerated by Festinger et al. (1956) are met by the True Word group. Therefore, if the theory as specified is valid, we should expect to observe an increase in fervor of belief, a greater openness to publicity, and strong attempts to proselyte upon their emergence from their shelters. This, of course, follows from the postulated need for the group members to reduce their dissonance and their inability to do this by any means other than by gaining new cognitions consonant with their belief.

Behavior of the Group Following the Disconfirmation

In the very early morning of August 16, the 103 "faithful" who had remained in the shelters for the full 42 days received the word to come out. At about 9 A.M. they held a joyous reunion in the church, led by their pastor, in which they were asked, "Did you have victory?" In unison came the reply, "Yes, praise the Lord!!!" Mrs. Shepard spoke of how their faith had not been shaken, "The Lord has brought the people closer to Him, there is not division, there's a fellowship here and we are the holiness people." Many other church members gave testimonies as to how their stay in the shelters had both strengthened their Christian fellowship and increased their belief.

The information concerning this first meeting was obtained from reporters who had been present. During the following week, the authors were able to speak with almost all of the members of the group, to attend their frequent church services, and to interview many of the members, including the leaders, Jameson, Blake, and Mrs. Shepard, quite intensively.

It is clear from our observations that the beliefs of the group remained intact. The group members did have a reinterpretation of the purpose of their stay in the shelters that served as an explanation for the failure of the prediction. They had discovered by looking back over all of their messages that it had never been stated that an attack was imminent; they had simply misinterpreted God's purposes. Really, God had just been using them to warn a world that was asleep, while at the same time He was testing their faith. They passed the test and thus proved themselves even more worthy to be among God's elect. We further discovered that they all continued to believe that an attack would come soon. Thus, the group members should be suffering from dissonance; the reinterpretation may have lessened it somewhat, by giving them some reason for having sat so long in the shelters. But their prediction had been shown to be wrong, and they still believed; they should, then, seek publicity and attempt to proselyte.

This did not occur; one must look very hard to uncover even the slightest indication that the members of the Church of the True Word wished to find new converts to their beliefs. The prayer meeting on the morning of August 16 was a golden opportunity if the group wanted to seek new believers. The press was there en masse, including several reporters, cameramen, and TV representatives. One newsman, who had kept in close touch with the group from the beginning of their stay in the shelters, did report that the group was a little more friendly to the press than formerly. Blake asked the press to print certain passages from the scriptures in their reports, and these passages, which speak of widespread destruction, are clearly intended as a warning to the world. Also, Mrs. Shepard, at this time and later, spoke favorably of all the free publicity they had gained for the Lord by the worldwide coverage of their activities. However, the group members were relatively indifferent to the attempts of Civil Defense officials to contact them and turned away curious tourists who asked to see their shelters. Furthermore, they made no immediate attempts to interest the townspeople in their church services. There is no indication from the behavior of the group when they first emerged or from our observations of them during the following week that they had any intentions of going out to seek new believers on a large scale.

DISCUSSION

The True Word group meets all of the criteria for a test of the theory as set forth in *When Prophecy Fails*, and yet their behavior following the disconfirmation does not conform to the expectations derived from that theory. Clearly, either the theory is wrong,[5] or it is incomplete in the sense that it specifies insufficiently the variables determining the predicted proselyting. We have two suggestions to make concerning differences between the True Word group and those previously studied that might have affected the differences in behavior that were observed.

The first difference that we wish to consider is that of the amount of social support present within the group. It is stated in *When Prophecy Fails* that one of the conditions necessary for proselyting to occur is the presence of social support for the believers. That is, a certain minimum amount of support is needed so that the individual believer may maintain his beliefs against the disconfirmation. But what might be the effect of additional amounts of social support? We would like to suggest that the more social support an individual receives above the minimum he needs to maintain his belief, the less need he will have to proselyte.

For this suggestion to be acceptable, it must be assumed, first, that there is some limit to the amount of support that is useful to an individual in his attempts to reduce his dissonance. For example, if only a few of your friends agree with a cherished belief of yours, you may be tempted to seek support by convincing others that you are right. On the other hand, if everyone with whom you associate agrees with you, you will feel very little need to go out and attempt to influence others in order to gain more support for your belief. Second, it must be assumed that a person will choose that means of reducing dissonance that is least likely to introduce new dissonance and most likely to reduce that which already exists. In the situation faced by the True Word group, and by other such groups, talking to other already convinced group members could not introduce new dissonance. In fact, interacting with others who had survived the same disconfirmation and who had emerged with their beliefs unshaken would be the best sort of support an individual could have. On the other hand, talking

[5] Leon Festinger, personal communication.

to the skeptical would be very likely to introduce new dissonance, since the person approached would probably reject one's attempts at influence and counter with arguments of his own. Thus, one would expect that, if at all possible, a person would choose to interact with those who agree with him rather than with the unbelievers.

In a group such as the True Word group with strong social support and a strongly shared belief system, the believer can turn to any other member for confirmation of his beliefs. Following our two assumptions, then, a member of such a group would first choose to talk to other members in his attempts to reduce his dissonance. If in this way he is able to garner new cognitions consonant with his belief up to the limit that he can use, he will then feel no need to seek further support by proselyting. On the other hand, a believer who is a member of a group such as the Lake City group, in which there is less support and more disagreement with regard to the belief system, would be much less likely to encounter sufficient support from his fellows. Thus, he may well have to resort to the otherwise less preferred means of gaining new consonant cognitions, that of proselyting.

It is easy to document the fact that the Lake City group did not provide social support to a degree that even approached that provided by the True Word group. First of all, the Lake City group was not well under way until about 5 months prior to the predicted date of the catastrophe; on the other hand, many members of the True Word group had worked together for several years. Further, the Lake City group had never lived as a community separated from the rest of the world as did the True Word group. This close association present in the True Word group should, one would think, foster a degree of trust in and understanding of the other members that far exceeded that which developed in the Lake City group.

With regard to shared beliefs, there was often disagreement among the members of the Lake City group. For example, the messages received by the two primary leaders of the group often contradicted each other. In contrast, the members of the True Word group were unanimous in their support of Mrs. Shepard, as far as we were able to observe. There was one leader, and one coherent set of beliefs shared by all of the group members. In conclusion, the Lake City group seems to have been characterized by only a minimal degree of social support, and we are suggesting that this degree of support was far from sufficient to reduce the dissonance suffered by the members as a result of the disconfirmation. As a result, the members, in search of further support, felt the need to proselyte. In contrast, the True Word group was very close and had a strongly shared belief system. Thus, they may well have had as much support within the group as they could utilize to reduce their dissonance and consequently felt no need to proselyte.

An interesting, although almost anecdotal, further piece of support for this suggestion comes from a very short article written by a Dutch psychologist, Van Peype (1960), in a Dutch newspaper. He briefly visited a group called the "Communita," who had gathered together near the top of Mont Blanc in expectation of a flood that would destroy the world on July 14. They had existed as a group for over 4 years, and many of the members had lived in the lodge on Mont Blanc for several months, separated from the other people living in the area. They had one prophet, one coherent body of beliefs, and, as far as Van Peype reports, a feeling of community and fellowship. When their prediction was disconfirmed, the leader announced to the waiting reporters, "You should be happy that we made that error. Our faith does not waver . . . Amen" (p. 3). He said no more to the assembled crowd and was reluctant to talk to Van Peype. The members had, then, not given up their belief and yet they showed no indications of a desire to proselyte. In the apparent degree of social support that was present, this group resembles the True Word group much more than it does the Lake City group, and their behavior following disconfirmation was essentially the same.

A second difference between the True Word group and the Lake City group that may have affected the amount of proselyting we observed lies in the amount of ridicule the groups received from the outside world. It would seem reasonable that if a group is receiving considerable ridicule from nonmembers, one way of reducing dissonance that would be apparent to them would be to convince these "unbelievers" that the group is right. If, however, the group is not receiving this sort of treatment from outsiders, this means of reducing dissonance would tend to be a great deal less salient to them. Furthermore, ridicule from others adds more dissonance to that which the group suffers from the disconfirmation. Thus, a very direct way of actually reducing part of the dissonance would be to eliminate the source of ridicule by converting the scoffers.

As far as we were able to determine, the True Word group received very little ridicule from townspeople and the press, considering the unusual step they had taken. The greatest amount

of censure the group received seemed to come from other evangelical churches rather than from people in general. In talking to the local townspeople we often encountered statements such as "Yes, we knew they were building bomb shelters. We believe in doing that, too." The Civil Defense officials in the area even presented the group with an award for "the service which they have performed for the public." After the group had left the shelters, the Mayor of the town was quoted in the newspaper as saying, "I sincerely hope no one ridicules them for their beliefs." Newspaper accounts of the group were in general factual and did not make fun of the group.

The treatment (Festinger et al., 1956) the Lake City group received was very different. Again, the news stories were generally factual and straightforward. But the headlines were cruel. In response to the announcement of the prediction, one paper headlined, "Tuesday—That Sinking Feeling," and another reported, "World Won't End, but Boy It Sure Will Shake" (p. 137). Columnists and editorial writers were equally unkind. Thus, since the Lake City group suffered more ridicule than did the True Word group, it might be supposed that it was easier for the Lake City group to see proselyting as an effective way of reducing the dissonance they had after the disconfirmation.

Evidence from the Mont Blanc group[6] gives somewhat equivocal support for this second suggestion. Van Peype reported that the Communita was very well thought of by the people in the town near their lodge, and was never ridiculed by them. However, they did receive some ridicule at the hands of reporters.

In conclusion, the True Word group, who had suffered a major disconfirmation of an important prediction, held to their beliefs and yet did not proselyte for them. This fact is in clear contradiction to expectation derived from the theory set forth in *When Prophecy Fails*. However, since dissonance theory has received considerable support in laboratory situations, it seems unlikely that it is completely wrong. Thus, we have assumed that the specification of the conditions that must obtain in the disconfirmation situation, in order that the predicted proselyting might occur, was insufficient. The two suggestions we have made for further conditions are that the group provide only minimal social support for its members and that the group receive ridicule from the outside world. It is, of course, impossible to know from the study of one group whether either of these has any relevance to proselyting. We can only say that there were clear differences on both these factors between the Lake City and True Word groups. We would like to suggest, then, that these two factors be considered and kept in mind by those investigating similar "doomsday groups" in the future.

[6] W. F. Van Peype, personal communication.

SUMMARY

On July 4, 1960, a group of 135 men, women, and children began a 42-day stay in underground shelters in response to a prophecy of widespread nuclear disaster. Since their situation bore marked similarities to that of the group studied in *When Prophecy Fails*, their reactions to the disconfirmation of their prediction were observed in order to test the theory set forth in that book. Although the group members clung to their belief in the face of disconfirmation, and even seemed to increase in fervor, they did not attempt to proselyte for their belief. This result is contrary to expectations derived from the observations and the theory presented by Festinger et al. (1956). Two variables, the degree of social support available within the group and the amount of ridicule the group received, are suggested as possibly having effected this difference in behavior.

REFERENCES

FESTINGER, L., RIECKEN, H., & SCHACHTER, S. *When prophecy fails*. Minneapolis: Univer. Minnesota Press, 1956.

VAN PEYPE, W. F. Nu de wereld op 14 Juli niet is vergaan. *Vrij Ned.*, July 30, 1960, 3.

(Received June 6, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 2, 142–144

EMOTIONAL AROUSAL AND STIMULUS DURATION AS DETERMINANTS OF STIMULUS SELECTION

JAMES M. MATTSON AND THOMAS NATSOULAS [1]

Wesleyan University

Blum (1954) has shown that, following presentation of a Blacky Picture in an emotional context, there is an increase in frequency of selecting that picture as *standing out* among four Blacky Pictures simultaneously flashed for .03 second. Under the same stimulus conditions, but with stimulus durations of .20 second, the frequency with which the picture is *located correctly* declines following the emotion arousing procedure. Blum predicted his results from the assumptions (*a*) that his arousal procedure would increase the sensitivity of the person to the picture involved, and (*b*) that the higher the level of awareness, the greater the degree to which defense mechanisms operate.

Dember (1960, p. 317) has criticized the experiment by Blum (1954) on two counts. First, the instructions asking the subject to pick the picture that stands out may lead him to choose the least clear because it differs from the rest. If this occurs, Blum's procedure cannot demonstrate an increase in sensitivity to the critical picture. Second, Blum used different instructions for the tests at each stimulus duration. When the stimuli were presented for a duration of .03 second, the subject had to indicate which of the four pictures *stood out;* when the stimuli were presented for a duration of .20 second, the subject was required to *locate* the critical picture in terms of its position (top, bottom, right, or left). The differential results found may, therefore, be attributed to something other than stimulus duration.

One can add a third criticism. The test series at .03 second was given first, immediately after the emotional manipulation. The test at .20 second followed.

The present experiment tests Blum's (1954) hypothesis under conditions of stimulus presentation somewhat different from the ones he used. No attempt is made, for example, to equate the two experiments with respect to size and illumination of the stimuli. The same assumption with regard to stimulus duration is made, however, that .03 second results in a low level of awareness and .20 second in a relatively higher level of awareness. In addition, the present experiment avoids the three criticisms outlined above by means of instructions to pick out the clearest picture, the same task for both stimulus durations, and two groups, one tested at .03 second and the other at .20 second right after the emotional arousal.

METHOD

Subjects. Twenty male freshmen at Wesleyan University participated for pay. All served without prior knowledge of the problem and procedure. They were assigned arbitrarily to one of two groups of 10 each. All subjects were tested individually by the same experimenter.

Apparatus and materials. Twenty-four 35-millimeter slides were prepared. These black and white positive transparencies were made by a positive reversal process with Panatomic X film. Four Blacky Pictures were used for each slide: Oral Sadism, Masturbation Guilt, Oral Eroticism, Identification Process (the same ones used by Blum, 1954). All possible arrangements of the four pictures in the top, bottom, right, and left positions of a diamond shape were used. The slides were projected onto a wall 8 feet from the Lafayette tachistoscope. The projector was directly to the right of the subject by about 4.5 feet. The slide projections gave an average light reading of .8 on the Weston scale, the background a reading of .1.

Procedure. The procedure, very similar to that of Blum (1954), can be described in terms of four successive steps.

1. Prior to subject's having any knowledge of the contents of the pictures, a baseline was determined for every subject by flashing the 24 slides once each at .03-second duration. The subject was required to indicate, by writing on a form, the position (top, bottom, left, or right) of the clearest picture. The slides were flashed in a corrected random order so that Masturbation Guilt never occurred at the same position on two successive slides. The same order was used in all parts of the experiment. The time between presentations of the slides within a series was about 15 seconds.

2. Every subject was shown two pictures from the Blacky series, Masturbation Guilt and Oral Sadism. Masturbation Guilt was placed in an emotional context: "Blacky can't help from licking his sexual parts even though he has been scolded for it." Oral Sadism was paired with a neutral description: "Blacky loves the free feeling of the out-of-doors and retrieves things like this belt or collar." Following the descriptions by the experimenter, the subject was asked

[1] Now at the University of Wisconsin.

to think of similar personal experiences for 60 seconds.

3. Group I again responded to the projections of the 24 slides at the .03-second duration. Group II responded to the same series under conditions of .20-second duration. For both groups the instructions were the same: to pick out the position at which the clearest picture appeared.

4. Group II was tested at the .03 level this time and Group I at the .20 level of stimulus duration. Again the clearest picture had to be selected.

The literal instructions were identical to those used by Blum (1954) except for small changes necessitated by the requirement throughout that the subject choose the position in which the clearest picture appears.

Design and analysis. Two difference scores, one for each stimulus duration, were calculated for every subject using the frequencies of his selections of Masturbation Guilt. The frequency of choosing the critical picture following the emotional arousal was corrected by subtracting the frequency found in the baseline determinations for that picture. For example, a positive difference score for the test at .03 second means that Masturbation Guilt was selected as clearest more frequently following emotional arousal than it had been selected before.

The scores for the stimulus durations were analyzed in a 2×2 Latin square as described by Lindquist (1953, pp. 273–281) for his "mixed" Type II design, order of test at a stimulus duration being the Latin square factor. A parallel procedure was followed in the analysis of choices of the "neutralized" picture of Oral Sadism.

RESULTS AND DISCUSSION

Table 1 presents the Means and *SD*s of the difference scores for Masturbation Guilt for the two groups under the two stimulus conditions. The table is a representation of the 2×2 design to which reference has been made.

Table 2 summarizes the analysis of variance of the difference scores for Masturbation Guilt. The *F* for stimulus durations, the only significant effect, has a probability of occurrence by chance of less than .005.

A parallel analysis (not summarized here) based on choices as clearest of Oral Sadism, the neutral picture, provided no significant effects.

TABLE 1

MEANS AND *SD*s OF DIFFERENCE SCORES FOR MASTURBATION GUILT

Group	Stimulus durations			
	.03 second		.20 second	
	M	*SD*	*M*	*SD*
I	+1.60	1.56	−0.10	2.21
II	+1.02	1.89	−1.30	2.76

TABLE 2

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES FOR MASTURBATION GUILT

Source	*df*	*MS*	*F*
Between subjects			
Groups	1	17.30	2.23*
Error	18	7.76	
Within subjects			
Durations	1	26.00	10.5**
Order	1	.50	0.20
Error	18	2.47	
Total	39		

* Significant at .10–.20 level.
** Significant at .001–.005 level.

Since the present experimental design confounds the interaction of order and duration with the main effect of groups, an evaluation of the interactive influence of order can only be made on the working assumption of no difference between the two groups other than that produced by differences in experimental treatment. Two *F* tests (simple randomized design) were performed, therefore, comparing the difference scores of the two groups under the .03 condition and under the .20 condition. Neither of them was statistically significant ($F = 2.94$ and 1.04, respectively, with $df = 1/18$ in both cases).

The hypothesis of Blum (1954) is supported by the present results despite certain departures from his conditions of stimulation and certain improvements on his procedure. Table 1 indicates a tendency for the emotional arousal procedure to make subjects more sensitive to the critical picture when the stimuli are presented at .03-second duration. On the other hand, a stimulus duration of .20 second results in fewer choices as clearest of the critical picture than were made prior to emotional arousal. Table 2 shows that this difference in behavior of the subjects under the two conditions is statistically significant.

Whether one tests with a particular stimulus duration immediately after the emotional manipulation or only after testing with the other duration does not seem to matter. It was expected that a reduction in effect would occur when the test was preceded by responses to the critical picture at the other duration. A strong suggestion of this, especially for the .20-second duration, is contained in Table 1. However, there is no statistical support for this conjecture, possibly because of large individual differences in rate of extinction of the emotional reaction.

SUMMARY

After receiving instructions which placed a Blacky Picture in an emotional context, two groups of 10 subjects were tested at two stimulus durations (.03 and .20 second) on the task of selecting, as the clearest, one from four Blacky Pictures (including the critical picture) presented simultaneously. The two groups differed in that Group I received the test at .03 second immediately after the emotional manipulation and the test with .20-second duration subsequently, while Group II received the test at .03 second only after having responded to the stimuli at .20-second duration. Thus, a 2×2 Latin square design (Lindquist's, 1953, "mixed" Type II) was used in which stimulus duration and experimental group were two factors and the Latin square factor was the order of test with each stimulus duration. The statistical analysis indicated that stimulus duration was the only effective factor. The emotional arousal procedure has the effect of increasing the frequency of choosing the critical picture as clearest at .03-second stimulus duration and of decreasing the frequency of choosing that picture at a stimulus duration of .20 second.

REFERENCES

BLUM, G. S. An experimental reunion of psychoanalytic theory with perceptual vigilance and defense. *J. abnorm. soc. Psychol.*, 1954, **49**, 94–98.

DEMBER, W. N. *The psychology of perception.* New York: Holt, 1960.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

(Received June 6, 1961)

An interdisciplinary journal for behavioral and social scientists who wish to keep abreast of research concerned with the problems of peace and international relations

The Journal of

conflict

RESOLUTION

A quarterly for research related to war and peace

Anaual subscription: $5.50
per copy: $2.00

Published by The Center for Research on Conflict Resolution
820 East Washington, University of Michigan, Ann Arbor, Michigan

"The new in tests. . . . . ."

Need an individual test which *quickly* provides a stable and reliable estimate of intelligence in 2 or 3 minutes per form? Has three forms? Does not depend on verbal production or subjective scoring? Can be used with the severely physically handicapped (even paralyzed), if they can signal yes-no? Handles two-year-olds and superior adults within the same short series of items and the same format? Has been painstakingly standardized? Builds rapport with intrinsically interesting and simple procedures based on multiple drawings?

Ammons and Ammons

NEW TEST!!

QUICK TEST (QT)

NEW TEST!!

Order Tester's Set including manual, card materials, keys, and 100 answer sheets ($8.00 complete).

•

For further information, write to

PSYCHOLOGICAL TEST SPECIALISTS
Box 1441
Missoula, Montana

"The new in tests. . . . . ."